Technische Universität München

Fakultät für Chemie

Lehrstuhl für Organische Chemie II

Development of Electrophiles and Chemoproteomic Methods for the Exploration of the Ligandability of Bacterial Proteomes

Dissertation zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften von

Patrick R. A. Zanon

München 2021

Technische Universität München

Fakultät für Chemie

Lehrstuhl für Organische Chemie II

Development of Electrophiles and Chemoproteomic Methods for the Exploration of the Ligandability of Bacterial Proteomes

Patrick R. A. Zanon

Vollständiger Abdruck der von der Fakultät für Chemie der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation

Vorsitzender:		Prof. Dr. Matthias J. Feige
Prüfer der Dissertation:	1.	Dr. Stephan M. Hacker
	2.	Prof. Dr. Cathleen Zeymer
	3.	Prof. Dr. Kathrin Lang

Die Dissertation wurde am 14.07.2021 bei der Technischen Universität München eingereicht und durch die Fakultät für Chemie am 04.08.2021 angenommen.

“I was taught that the way of progress was neither swift nor easy.”

Marie Skłodowska-Curie

Danksagung

An erster Stelle möchte ich Dr. Stephan Hacker danken für vier Jahre voller wissenschaftlicher Diskussionen auf einer sehr kollegialen Ebene und ein stets offenes Ohr für Fragen und Probleme innerhalb und außerhalb des Labors. Es ist alles andere als selbstverständlich, dass ein junger Gruppenleiter den täglichen Druck des akademischen Karriereweges nicht überträgt, sondern in allen Situationen stets ruhig und gelassen mit seinen Mitarbeitern umgeht. Durch die gemeinsame Zeit habe ich deine Führungsqualitäten, dein Fachwissen und deine Begeisterung für Wissenschaft im Allgemeinen sehr schätzen gelernt. Ich bin ausgesprochen dankbar für das Vertrauen, dass du mir von Anfang an entgegengebracht hast, deine außerordentliche Unterstützung und die herausragende Zusammenarbeit. Ich bin mir sicher, dass du auch die nächsten beruflichen und familiären Schritte in Leiden hervorragend meistern wirst und wünsche dir, Rica und Georg alles Gute!

Ein großer Dank geht an die weiteren Mitglieder meiner Prüfungskommission, Prof. Dr. Matthias J. Feige, Prof. Dr. Cathleen Zeymer und Prof. Dr. Kathrin Lang. Vielen lieben Dank an der Stelle auch an Patrick Allihn, Ines Hübner, Till Reinhardt, Angela Weigert-Muñoz, Michael Zollo und Stephan Hacker für das Korrekturlesen dieser Arbeit!

Ich bin außerdem dankbar für die erfolgreichen Kollaborationen außerhalb des Lehrstuhls mit Dr. Fengchao Yu, Prof. Dr. Alexey I. Nesvizhskii, Prof. Dr. Patricia Zhang Musacchio, Kristina Krauskopf, Dr. Marko Cigler, Prof. Dr. Kathrin Lang, Prof. Dr. Christopher J. Chang, Prof. Dr. F. Dean Toste, Dr. Sébastien Britton, Dr. Marta Bogacz, Dr. Annika Wagner und Prof. Dr. Ute Hellmich.

Über die Jahre durfte ich die Hacker-Labore mit einigen Kollegen teilen. Vielen lieben Dank, Lisa Lewald, für die gute Laune, die du jeden Tag ins Labor gebracht hast! Ein großes Dankeschön auch an das bayrische Unikat Michael Zollo, es war immer eine Gaudi mit dir! Unserem Langzeitpraktikanten, Thomas Maher, danke ich für seinen resoluten Optimismus und unbeschwerte Natur. Kathrin Bach und Bert Beerkens möchte ich für die gute Zusammenarbeit auf dem Tetrazol-Projekt danken. Ein großes Dankeschön gilt auch Dario Mrdović für seine nüchterne und offene Art und Patrick Raunft für seine Hartnäckigkeit bei der Synthese. Natürlich möchte ich auch meinen Praktikanten Clara Hipp, Maurits Brinkman, Max Müller, Lisa Schlor, Philipp Weingarten, Simon Kollmannsberger und Maximilian Iglhaut danken, die mich tatkräftig auf verschiedenen Projekten unterstützt haben.

Ohne das Zutun von Prof. Dr. Stephan Sieber wären all diese Begegnungen vermutlich nicht zustande gekommen. Ich bin sehr dankbar dafür, dass er damals den Kontakt mit Stephan

Hacker hergestellt hat und mich als dessen ersten Doktoranden empfohlen hat. Als Juniorgruppe wurden wir sehr herzlich in Lehrstuhl OCII aufgenommen, wo es uns fachlich sowie menschlich an nichts gefehlt an. Für die gute Zeit im Labor möchte ich allen in der Gruppe danken! Vielen Dank, Josef Braun für die Gaudi im Labor und beim JCF, Konstantin Eckel für die Gesellschaft bei langen Laborabenden, Dr. Carolin Gleißner für den ein oder anderen Aperol oder Mexikaner zu viel, Dr. Thomas Gronauer für seinen flachen Humor, Dietrich Mostert für seine erfrischende Art, Martin Pfanzelt für erbitterte Duelle im Squashcourt, Dr. Theresa Rauh für die vielen Morgen in der Boulderhalle und Dr. Stuart Ruddell für die Hilfe bei Lebensentscheidungen. Außerdem danke ich Dr. Nina Bach, Dr. Dóra Balogh, Katja Bäuml, Davide Boldini, Andrei Bubeneck, Jan-Niklas Dienemann, Michaela Fiedler, Alexandra Geißler, Martin Köllen, Dr. Franziska Mandl, Dominik Schum, Markus Schwarz, Dr. Isabel Wilkinson sowie den Alumni Ramona Absmeier, Tobias Becker, Dr. Barbara Eyermann, Dr. Christian Fetzer, Dr. Anja Fux, Dr. Mathias Hackl, Dr. Wolfgang Heydenreuter, Dr. Volker Kirsch, Dr. Pavel Kielkowski, Dr. Vadim Korotkov, Dr. Elena Kunold, Dr. Markus Lakemeyer, Dr. Philipp Le, Dr. Kyu Myung Lee, Dr. Igor Pavlović und Dr. Matthias Stahl. Vielen Dank auch an Mona Wolff, Christina Brumer und Barbara Seibold, ohne die der ganze Laden vermutlich schon lange in sich zusammengefallen wäre.

Mit ein paar Lehrstuhlmitarbeitern bin ich besonders oft noch am Feierabend in der Kaffeeküche versumpft und diesem harten Kern möchte ich ganz besonders danken. Herzlichen Dank, Patrick Allihn, dafür, dass du dich mehr über meine Erfolge gefreut hast als ich selbst und deine unermüdliche Ausdauer in der Forschung, beim Sport und am Glas. Vielen lieben Dank, Dr. Jonas Drechsel, für deine abgeklärte Art und die fachmännische Installation meiner Deckenleuchten im Rahmen der Alte Heide-Nachbarschaftshilfe. Vielen lieben Dank, Dr. Ines Hübner, für dein scharfes Auge beim Erstellen von Grafiken, deine Fürsorge in Notsituationen und Unmengen Blödeleien. Danke, Dr. Robert Macsics, für die teils willkürlichen Auszüge aus dem Brockhaus und deine Skepsis *in puncto* Limogehalt fremder Biere. Vielen Dank, Till Reinhardt, für dein strahlendes Gemüt, dem selbst der Laboralltag nur selten etwas anhaben kann, und Bouldersessions mit Burgern und Bier. Herzlichen Dank, Angela-Maria Weigert Muñoz, für einen ungewollten Undercut und dafür, dass ich um 22:30 Uhr immer wieder etwas zu schmunzeln habe.

Des Weiteren will ich mich herzlichst bei der ganzen Damn-Gruppe bedanken, die mich größtenteils schon durch das Studium und seitdem auch durch viele Höhen und manche Tiefen begleitet haben. Vielen lieben Dank, Brigita Bratic, dafür, dass du mich in die kroatische Fankurve aufgenommen hast und für jeden Spaß zu haben bist. Herzlichen Dank, Dr. Annika

Haugeneder, für geteilte Musikvorlieben und deine übertrieben lässige Art. Vielen lieben Dank, Dr. Sebastian Hölzl, für viele schwachsinnige Geschäftsideen, das Etablieren von Traditionen und Gründung der Muskeltiere. Vielen Dank, Dr. Fabian Hörmann, für die gemeinsamen Praktika und weil jeder Tag mit dir „best day ever“ ist. Herzlichen Dank, Lara Milakovic, für viel Rückhalt und unvergessliche Muskeltier-Aktionen. Fettes Merci, Dr. Jens Oberkofler für die coolsten Dance Moves und Freundschaft seit der zweiten Vorlesungswoche des Bachelors. Vielen Dank, Dr. Lorenz Pardatscher, für die Personifizierung Südtiroler Lässigkeit und sagenhafter (wenn auch teils erzwungener) Gastfreundschaft. Herzlichen Dank, Dr. Christina Schneider, für die Zeit beim JCF und deine ruhige und zielstrebige Lebensweise, von der man sich definitiv eine Scheibe abschneiden kann. Vielen lieben Dank, Anna Trebo für deine direkte Natur und deinen trockenen Witz. Vielen Dank, Gerhard Vorlauff für Ernährungstipps und weil du einfach bester Mann bist. Danke, Dr. Kathrin Weger, für viel Planung in allen Lebensbereichen und dafür, dass du mir Schafkopf beigebracht hast. Dr. Michael Weger für die niederbayrische Unterstützung und die langjährige Leitung des besten Clubs in Garching.

Vielen Dank gilt auch meinen Schulfreunden, die mich zum Teil schon knapp zwei Jahrzehnte begleiten. Allen voran möchte ich der Bosca-Crew bestehend aus Simon Spitzer und Christian Zellmer danken für viele philosophische und ausufernde Abende.

Länger hat es ansonsten (gezwungenermaßen) bisher nur meine Familie mit mir ausgehalten. Ich danke meinen Eltern Alexander Zanon und Monika Kaschny für all die Möglichkeiten, die sie mir gegeben haben. Meinen Brüdern Alexander und Luca Zanon danke ich herzlichst für Ihren Rückhalt und viele Jahre voll Jux und Tollerei.

Ausgesprochen dankbar bin ich Katherine Duncker für ihre verständnis- und liebevolle Unterstützung über den großen Teich hinweg.

Abstract

The discovery of novel antibiotic targets and lead structures is essential to combat the antibiotic resistance crisis. Competitive, residue-specific chemoproteomic methods enable the efficient discovery of binding sites of covalent ligands across the proteome and thus are valuable tools for drug development. In this work, a new chemoproteomic platform is described that utilizes isotopically labeled desthiobiotin azide (isoDTB) tags, which are easy to synthesize and compatible with a wide variety of experimental conditions. To prioritize compounds for chemoproteomic target profiling using the isoDTB tags, phenotypic screening of a cysteine-directed covalent fragment library for activity against the pathogenic bacterium *Staphylococcus aureus* was performed. The most potent covalent ligands were selected for target deconvolution, which streamlined the proteome-wide identification of ligandable cysteines and molecular lead structures to address them.

While such a competitive approach avoids the need for modification of the hit compounds with a reporter tag, this method of target discovery relies on the use of a broadly reactive alkyne probe. To enable the development of targeted covalent inhibitors addressing nucleophilic residues other than cysteine, selective probes for the respective amino acids are required. 2,5-Disubstituted tetrazoles were established in this thesis to monitor aspartates and glutamates proteome-wide. As photoactivated probes, they are readily applicable to study these amino acids even in living cells. Moreover, the mechanism of their high selectivity for these residues was investigated on a small molecule-level. As only few covalent ligands targeting carboxylic acid residues have been reported so far, these probes will foster their development, with hydrazonoyl chlorides being a promising new compound class for covalent inhibition.

Although a wide variety of different chemistries has been described to modify nucleophilic amino acid residues in proteins, only few studies have been carried out in a proteomic context and a direct comparison of different reactive groups was still entirely missing. To provide such a comparative analysis of the utility of different electrophiles for covalent inhibitor development and broadly reactive probe design, more than 50 probes were synthesized and tested for their proteome-wide reactivity and selectivity. For this purpose, the bioinformatic platform FragPipe was tailored for the robust and unbiased analysis of the chemoproteomic data. As a result, reactive probes for the proteome-wide profiling of nine different amino acids as well as the *N*-terminus were verified or newly identified.

Zusammenfassung

Die Entdeckung neuer antibiotischer Zielproteine und Leitstrukturen ist essenziell für die Bekämpfung der Antibiotikakrise. Kompetitive, positionsaufgelöste chemoproteomische Methoden ermöglichen die effiziente proteomweite Identifizierung der Bindestellen kovalenter Liganden und sind daher wertvolle Werkzeuge für die Wirkstoffentwicklung. Hier wird eine neue chemoproteomische Plattform beschrieben, die isotopenmarkierte Desthiobiotinazid (isoDTB)-Tags einsetzt, welche leicht zu synthetisieren und mit vielen verschiedenen experimentellen Bedingungen kompatibel sind. Zur Priorisierung von Verbindungen für die Identifizierung ihrer Zielproteine mit den isoDTB-Tags wurde ein phänotypisches Screening einer Cystein-gerichteten kovalenten Fragment-Bibliothek auf antibakterielle Aktivität gegen das pathogene Bakterium *Staphylococcus aureus* durchgeführt. Die wirksamsten Liganden wurden für die Ermittlung ihrer Zielstrukturen ausgewählt, wodurch sowohl adressierbare Cysteine als auch erste Liganden, um diese weiter zu untersuchen, identifiziert wurden.

Während die Notwendigkeit der Modifikation eines Hits mit einem Reporter-Tag in einem solchen kompetitiven Ansatz umgangen wird, ist diese Methode abhängig von der Verfügbarkeit einer geeigneten reaktiven Alkinsonde. Für die Entwicklung kovalenter Inhibitoren, die andere nukleophile Aminosäuren als Cystein adressieren, werden daher selektive Sonden für die entsprechenden Aminosäuren benötigt. 2,5-Disubstituierte Tetrazole wurden in dieser Arbeit zur proteomweiten Untersuchung von Aspartaten und Glutamaten etabliert und können als lichtaktivierbare Sonden sogar in lebenden Zellen angewandt werden. Außerdem wurde der Mechanismus untersucht, der der hohen Selektivität für diese Aminosäuren zugrunde liegt. Nachdem bisher nur wenige kovalente Liganden für Carbonsäure-Reste beschrieben wurden, werden diese Sonden ihre weitere Entwicklung fördern, wobei Hydrazonoylchloride eine vielversprechende neue Klasse kovalenter Liganden darstellen.

Trotz der großen Vielzahl chemischer Methoden, die zur Modifikation nukleophiler Aminosäurereste in Proteinen beschrieben wurden, wurden nur wenige Studien in einem kompletten Proteom durchgeführt und ein direkter Vergleich verschiedener reaktiver Gruppen fehlte gänzlich. Um die Eignung verschiedener Elektrophile für die Entwicklung kovalenter Inhibitoren und reaktiver Sonden zu untersuchen, wurden hier mehr als 50 Sonden synthetisiert und hinsichtlich ihrer proteomweiten Reaktivität und Selektivität analysiert. Hierzu wurde die Software FragPipe für die robuste und unvoreingenommene Analyse der chemoproteomischen Daten optimiert. Auf diese Weise wurden reaktive Sonden für die proteomweite Untersuchung von neun verschiedenen Aminosäuren und des *N*-Terminus verifiziert oder neu etabliert.

Introductory Remarks

This dissertation was completed between August 2017 and July 2021 under the supervision of Dr. Stephan M. Hacker at the Chair of Organic Chemistry II at the Technical University of Munich.

Parts of this thesis have been published:

Isotopically Labeled Desthiobiotin Azide (isoDTB) Tags Enable Global Profiling of the Bacterial Cysteinome
Patrick R. A. Zanon, Lisa Lewald, Stephan M. Hacker
Angewandte Chemie International Edition **2020**, *59*, 2829-2836.

Light-Activatable, 2,5-Disubstituted Tetrazoles for the Proteome-Wide Profiling of Aspartates and Glutamates in Living Bacteria
Kathrin Bach*, Bert L. H. Beerkens*, Patrick R. A. Zanon*, Stephan M. Hacker
ACS Central Science **2020**, *6*, 546-554.
(* these authors contributed equally)

Profiling the Proteome-Wide Selectivity of Diverse Electrophiles
Patrick R. A. Zanon, Fengchao Yu, Patricia Z. Musacchio, Lisa Lewald, Michael Zollo, Kristina Krauskopf, Dario Mrdović, Patrick Raunft, Thomas E. Maher, Marko Cigler, Christopher J. Chang, Kathrin Lang, F. Dean Toste, Alexey I. Nesvizhskii, Stephan M. Hacker
ChemRxiv **2021**. DOI: 10.26434/chemrxiv.14186561 – pending peer review

Parts of this thesis have been presented at conferences:

EMBL Conference - Expanding the Druggable Proteome with Chemical Biology 2020
5th – 7th February 2020, Heidelberg, Germany
Poster presentation, poster prize winner

RSC Chemical Biology and Bio-Organic Group Postgraduate Symposium 2020
10th – 11th September 2020, virtual conference
Poster presentation

Nature Conference Translational Chemical Biology 2020
5th – 6th October 2020, virtual conference
Flash talk

London Proteomics Discussion Group Webinar Series
18th June 2021, virtual webinar
Invited talk

Contents

I. Introduction

1. Antibiotic Resistance

Emergence of Resistance

Since the discovery and commercialization of antibiotics in the first half of the 20th century, they have immensely contributed to raising the global life expectancy and enabled many advances in modern medicine.[1] With increasing numbers of antibiotic-resistant bacterial strains, however, this progress is being undermined.[2]

Rapid mutations allow bacteria to adapt to their environment and compete with other microbes. While resistance development is thus a natural evolutionary trait, careless and wrongful use of antibiotics has accelerated this process.[2] Out of all antibiotic prescriptions in ambulatory care, more than 30% and 50% were estimated to be inappropriate in the USA (2008-2011)[3] and China (2014-2018),[4] respectively. The by far largest area of use of antibiotics, however, is not in human medicine but livestock production.[5] Here, antibiotics are often used for nontherapeutic purposes such as substitution of hygiene and promotion of animal growth.[5] Such a continuous, selective pressure facilitates the emergence of bacterial resistance, which can spread into the environment and human pathogens.[6]

Reports of resistant strains against all marketed antibiotics paint a bleak picture and advocate investment in countermeasures.[2] Among the most pressing concerns is the loss of efficacy of antibiotics like carbapenems and vancomycin, which have been kept as a “last line of defense” for the most serious infections.[7] Even though this stewardship strategy, to only prescribe new antibiotics if everything else has already failed, is important to curb resistance development, it has also counterintuitively contributed to the current antibiotic crisis. From a business perspective, the arduous and expensive development of antibiotics necessitates sufficient revenue to be generated from a new drug. Short treatment durations and a strongly limited number of prescriptions for antibiotics diminish profit, rendering the market very unattractive.[8] Consequently, pharma companies have largely abandoned the research and development of antibiotics, with only very few new classes of antibiotics in clinical trials[9] and the vast majority of preclinical projects being carried out in small and medium-sized enterprises.[10]

A Silent Pandemic

As a result of the discrepancy between the need for and the lack of innovation,[11] antimicrobial resistance is considered one of the greatest present challenges of humanity.[12] The dangers of infectious diseases are more apparent than ever in our lifetime – SARS-CoV-2 is taking a

harrowing toll on every part of our heavily interconnected world. More than four million people have died and the long-term effects on survivors are still only beginning to be understood.[13] Even without being infected themselves, many more suffer from the impacts on mental health and the economy.[14] While antibiotics can play a key role in the treatment of bacterial coinfections in respiratory viral diseases,[15] there are concerns that their widespread preemptive use in COVID-19 patients will lead to a surge in antibiotic resistance.[16,17]

Antimicrobial resistance currently already claims more than an estimated 700,000 people annually[18] but is predicted to overtake cancer as the leading cause of death by 2050 and claim the lives of up to 10 million people every year.[19] Furthermore, this crisis is projected to force 28 million people into extreme poverty and may cut 3.8% of the annual global gross domestic product.[20] Taken together with the global spread of resistance genes,[21,22] it is clear that global collaboration for prevention and mitigation of outbreaks of resistant strains is imperative and that actions from various angles are required.[2,23]

Fighting Antibiotic Resistance

Although poor general access to antibiotics remains to be a driver in mortality in low-income countries,[1] antibiotic resistance is also prevalent there.[24] Many deaths could be avoided by making second- and third-line antibiotics more available,[25] which, however, has to go hand in hand with proper antibiotic stewardship. This conundrum of making antibiotics more available, while preventing a quick rise and spread of resistance warrants policies that get continuously reviewed and adjusted to maximize therapeutic benefit across countries.[2]

In order to stimulate the antibiotics industry, various strategies have been proposed and to some extent also implemented.[23] So-called push incentives are currently most common and reduce the cost of development for example through direct funding and reduced taxes and thus also cover unsuccessful projects. However, they do not guarantee any profit to be made from an approved drug. Pull mechanisms on the other hand reward milestones during the development and increase revenue for example through extended patents.[23] While it is unclear, which measures will prove most effective without endangering access and stewardship to both new and current antibiotics, it is of utmost importance to foster innovation in the field.

Maximizing The Value of Current Antibiotics

The application of antibacterial drugs could be largely improved by the identification of the pathogen and its resistance profile. This approach allows the informed selection of an

appropriate and narrower-spectrum antibiotic to benefit the patient outcome and to monitor and minimize resistance formation and spread.[26,27] Current technologies however are limited by costs and access to laboratories, particularly in low-income countries.[28]

Another promising approach to extend the clinical utility of antibiotics is to use them in combination with other drugs. By pairing the antibiotic with an inhibitor of the resistance mechanism, therapeutic sensitivity can be restored, as in the case of β-lactams and β-lactamase inhibitors. This strategy, however, is itself prone to be circumvented by bacterial evolution.[29] Recently, selection inversion has emerged as a promising concept to impose an evolutionary disadvantage on the resistant strain for example if resistance to one drug results in a heightened sensitivity to the other. Although this approach bears high potential as it goes beyond mere restoration of susceptibility, identifying suitable combinations is non-trivial.[29]

Discovery of New Antibiotics

In the so-called golden age of antibiotics, natural product screens have led to the discovery of most of all current classes of antibiotics.[30] Even though this approach has been far less fruitful in the decades that followed, advances in cultivating bacterial communities and the analysis of biosynthetic gene clusters have enabled new discoveries. One notable example is teixobactin (Figure 1a),[31] against which no resistances have been reported so far. While this outstanding characteristic makes it a formidable first-in-class addition to the antibiotic arsenal, its antibiotic activity – like most other antibiotics in the clinical pipeline[7] – is limited to Gram-positive bacteria.[31]

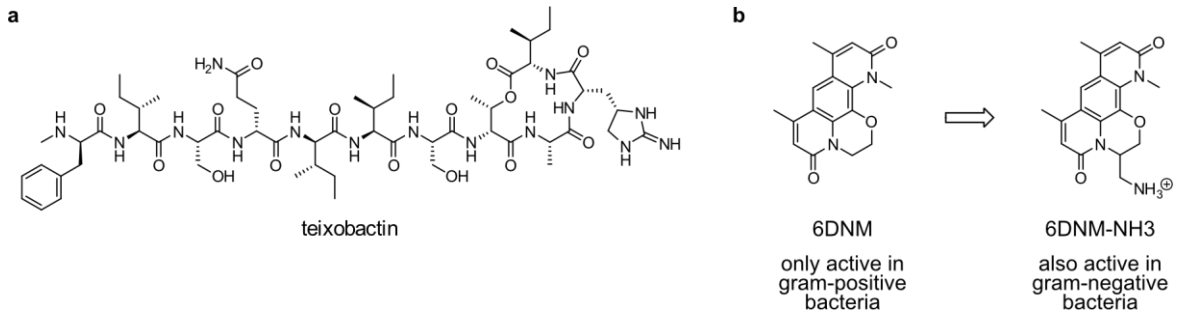


Figure 1 | a) Structure of teixobactin.[31] **b)** Introduction of a primary amine into the structure of the deoxynybomycin derivative 6DNM expands its antibiotic activity from only Gram-positive to Gram-negative strains.[32]

Most of the pathogens prioritized by the World Health Organization (WHO) are Gram-negative bacteria,[7] as their double-membrane cell envelope[33] and efflux pump systems[34] exacerbate in-cell accumulation of antibiotics. Only recently it was found that highly rigid, flat molecules bearing a positive charge show enhanced uptake into these organisms. Remarkably, antibiotics

that are only active in Gram-positive strains can gain activity in Gram-negative bacteria through modification of the molecular structure with a positively charged functional group (see Figure 1b for an example).[32,35,36]

Although extending the potency of existing classes of antibiotics to Gram-negative bacteria provides an obvious benefit, the mechanisms of action of all current antibiotics are limited to few physiological processes, namely membrane or cell wall disruption and inhibition of synthesis of proteins, folate, or nucleic acids.[37] Although bacteria possess hundreds of essential genes,[38] there are only very few antibiotics in development that engage new antibacterial targets.[9] Accordingly, efforts towards the discovery of novel therapeutics addressing an expanded target space are crucial for a sustained response to bacterial pathogens.[39] While other strategies to combat bacterial infections like antivirulence[40] and phage therapy[41] exist, inhibition of new antibacterial targets will undoubtedly make important contributions in the fight against antibiotic resistance.

2. Covalent Inhibitors

Covalent Inhibitors are Privileged Antibiotics

β-Lactams are the largest class of antibiotics listed as essential medicines by the WHO.[42] They, as well as the essential reserve antibiotic fosfomycin, (Figure 2) act through the formation of a covalent bond with their target proteins.[43,44] Despite this prevalence of covalent modes of action in the most valuable antibacterial agents, industrial research and development has long shied away from developing covalent drugs due to concerns that the reactivity will inevitably also lead to non-specific modifications and hence toxicity.[45] While the screening of synthetic libraries for antibacterial activity has been mostly unfruitful in the past decades,[39] covalent inhibitors have been considered liabilities at the time and now offer unique avenues towards novel antibiotics.

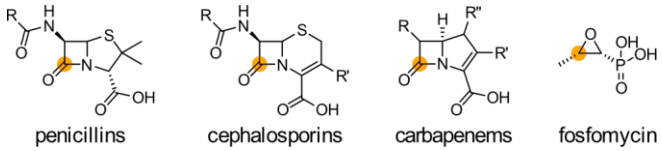


Figure 2 | Structures of penicillin, cephalosporin, and carbapenem subclasses of β-lactam antibiotics and fosfomycin. The site of nucleophilic attack is indicated by an orange circle.

Properties of Covalent Ligands

The binding of covalent ligands to proteins can be described as the sequential formation of a non-covalent complex and a covalent bond (Figure 3a).[46] Accordingly, covalency increases the ligand binding beyond the equilibrium of non-covalent interaction (defined by the dissociation constant K_i), especially for irreversible covalent bond formation ($k_{-2} = 0$).[46] In that case, k_2 is often referred to as k_{inact} and the overall ligand potency is expressed by the time-independent, apparent second-order rate constant k_{inact}/K_i (Figure 3b).[46] While the extent of the effect varies for reversible covalent ligands, the formed covalent bond increases the biochemical efficiency of the ligand.

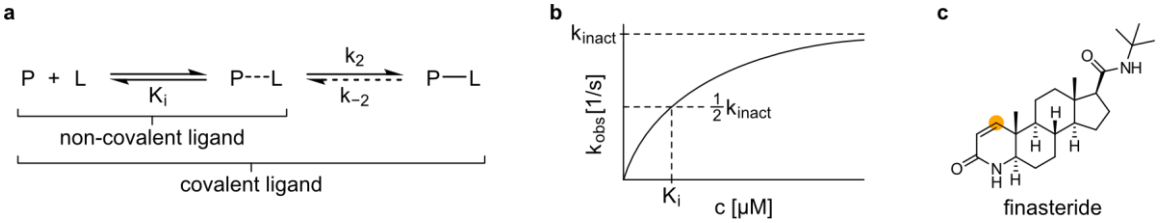


Figure 3 | a) (Non-)covalent protein-ligand interaction and kinetic descriptors. P: protein. L: ligand. **b)** Schematic representation of k_{inact}/K_i. c: inhibitor concentration, k_{obs}: observed inhibition rate constant. **c)** Structure of finasteride. The site of nucleophilic attack is indicated by an orange circle.

The binding kinetics of covalent inhibitors also have important implications for the pharmacological properties, as their increased target residence time results in an extended physiological effect even after clearance of the drug. Additionally, a swift clearance of unbound covalent ligand can result in selectivity as it reduces off-target binding and associated toxicity.[45] In contrast, non-covalent ligands are dependent on prolonged high concentrations of unbound ligand to maintain target occupancy.[45] The covalent antiandrogen finasteride (Figure 3c), for example, has a dissociation half-life from its target of 31 days[47] and the physiological effect of a single dose lasts for up to four days.[48] Such a sustained therapeutic effect increases dose intervals and thus patient compliance to medication.[49] However, this uncoupling of pharmacodynamics from pharmacokinetics through prolonged residence time is limited if the inhibited protein is rapidly resynthesized *in vivo*.[45,50]

Another strength of covalent inhibitors is that they can be able to achieve high selectivity between targets with high sequence homology (*e.g.* kinases), which usually poses a significant challenge to drug development. Here, covalency provides a distinct advantage, since an ideally positioned covalent binding element can react with non-conserved nucleophilic residues in the binding pocket.[45,51]

Targeted Covalent Inhibitors

Covalent inhibitors that engage nucleophilic protein residues without leveraging an enzymatic mechanism are also commonly referred to as targeted covalent inhibitors (TCIs). In contrast, mechanism-based inhibitors like penicillins[52] are substrate analogs that react with the catalytic residue and inhibit it by covalent modification.[53]

Large, shallow binding pockets like protein-protein interfaces are hard to address with non-covalent interactions alone due to high surface exposure and limited binding surface area.[54] As a result of the stabilized protein-ligand complex, TCIs bear great potential to engage such sites.[55] Intriguingly, TCIs might also be less prone to resistance formation as mutated sites in the binding pocket will primarily affect the first, non-covalent step but, as long as the nucleophilic residue is not mutated, not the covalent binding. Accordingly, effective target occupancy can still be achieved in spite of less favorable non-covalent interactions.[45] In the case of the cancer target EGFR, TCIs were able to overcome a gatekeeper mutation, which had led to resistance against non-covalent drugs.[56] In some of the patients, however, mutation of the nucleophilic cysteine to a much less reactive serine has diminished covalent bond formation and thus drug efficacy.[57]

Approved Covalent Drugs

Highly reactive nitrogen mustards are unspecific alkylators of DNA and proteins and accordingly exhibit high toxicity (Figure 4a). While they find clinical application as chemotherapeutics in severe forms of cancer,[58] the serious adverse effects stress the importance of tempered reactivity of the covalent motif. More than 50 covalent drugs have been approved also for less dire conditions as they achieve appropriate reactivity, target specificity, and pharmacokinetics.[59] Besides cancer (*e.g.* afatinib,[60] sotorasib[61]) and infectious diseases (*vide supra*), cardiovascular (*e.g.* aspirin[62]) and central nervous system conditions (*e.g.* dimethyl fumarate[63]) are the main indications of covalent drugs, evidencing their broad applicability (Figure 4b).[59]

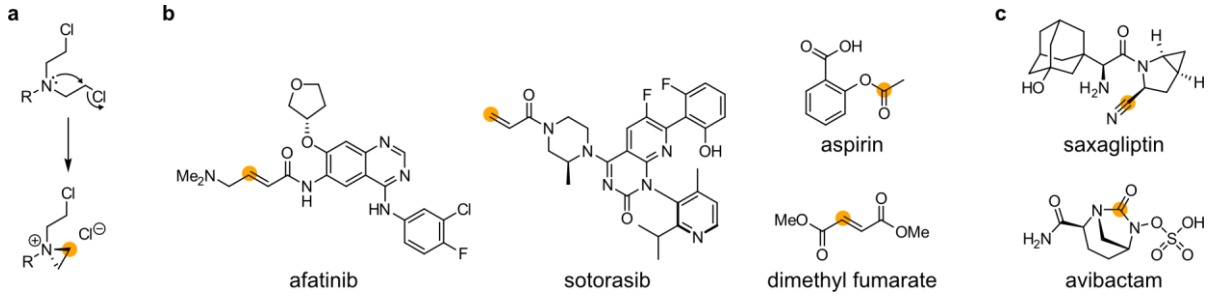


Figure 4 | a) Spontaneous formation of reactive alkylating species from nitrogen mustards.[58] **b)** Structures of a selection of approved covalent drugs.[60-63] **c)** Structures of selected reversible covalent drugs. The site of nucleophilic attack is indicated by an orange circle.

The FDA approval of sotorasib (Figure 4b) in May 2021 marks the most recent achievement of the field of TCIs.[64] Even though KRAS was among the first oncogenes to be identified almost 40 years ago, all efforts to produce an effective and safe drug had failed until recently, resulting in the KRAS protein being considered "undruggable".[65] Non-covalent approaches were not able to outcompete the tightly bound endogenous ligands GDP and GTP in the only tractable binding site of KRAS.[61,65] However, the common malignant mutation G12C in a shallow allosteric pocket can be leveraged for covalent inhibition[66] and sotorasib is the first drug that successfully engages this therapeutically valuable target.[61]

While the majority of approved covalent drugs are considered to react irreversibly, reversible covalent drugs ($k_{-2} \neq 0$) can combine the advantages of increased potency through covalency and thermodynamic equilibration to avoid off-target effects.[67] Two such examples, the anti-diabetic drug saxagliptin[68] and the β-lactamase inhibitor avibactam[69] (Figure 4c) reversibly react with the catalytic serines of their target proteins.

Cysteine-Reactive Electrophiles

Cysteines and catalytic serines constitute the vast majority of amino acid residues targeted by covalent drugs.[70] While non-catalytic serines, in general, are weak nucleophiles, cysteine bears the intrinsically most nucleophilic sidechain of all proteinogenic amino acids and is amenable to covalent modification.[70] As a result, a plethora of electrophilic groups has been incorporated in TCIs to target cysteines, with acrylamides being the most common.[71] The reactivity of a simple acrylamide (**1**, *e.g.* sotorasib) can be significantly reduced by β-methyl- (**2**) or α-fluoro-substitution (**3**) (Figure 5a).[72] Interestingly, the introduction of a nitrile in the α-position renders the reaction with cysteines highly reversible and thus offers a simple way to turn an irreversible TCI into a reversible one.[73] Other *Michael* acceptors like propiolamides[74] (**5**) and vinylsulfonamides[66] (**6**) also have been used as cysteine-directed electrophiles.

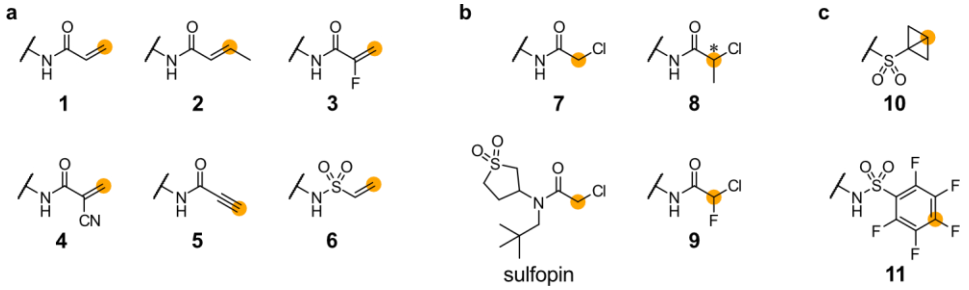


Figure 5 | Selection of *Michael* acceptors (**a**), chloroacetamides (**b**), and other electrophiles (**c**) used in cysteine-targeting covalent inhibitors. The site of nucleophilic attack is indicated by an orange circle.

Although chloroacetamides (**7**) are often regarded as too reactive for covalent drugs, tempered reaction rates similar to acrylamides can also be achieved through the appended motif (Figure 5b).[75] Recently, sulfopin has been shown to be an effective and safe inhibitor of the peptidyl-prolyl isomerase in *in vivo* studies.[76] In addition to substitution of the amide, α-substitution allows further reactivity tuning. Methylation (**8**) not only significantly reduces the reactivity but also introduces a stereocenter, influencing the orientation of the reactive group towards protein nucleophiles resulting in distinct target profiles of two enantiomeric TCIs.[77] Chlorofluoroacetamides (**9**) show reduced reactivity and the formed cysteine adducts are prone to hydrolysis when they are solvent-accessible, which could serve as a potential mechanism for reduced off-target modification.[78] Strained electrophiles like bicyclo[1.1.0]butane sulfones[79] (**10**) and electrophilic aromatic systems like pentafluorophenylsulfonamides[80] (**11**) are additional examples of a wide variety of cysteine-reactive moieties (Figure 5c).[71]

Targeting Amino Acids Beyond Cysteine

Even though cysteine offers many possibilities for TCI development, it is a rare amino acid[81] (Figure 6a) and many proteins and protein sites do not have an appropriately reactive cysteine present. By focusing research efforts on cysteine-targeting TCIs, the potential target space is also limited to a small subset of the proteome. In order to fully exploit the advantages of TCIs and develop drugs with novel biological modes of action, selective modification of other nucleophilic side chains (Figure 6b) must be realized.

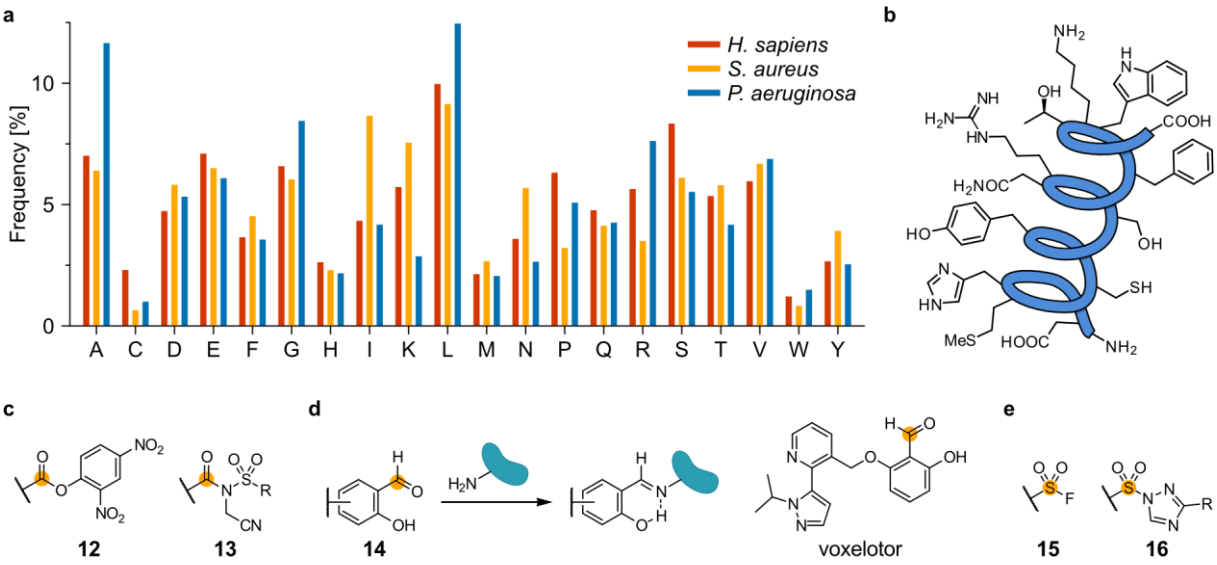


Figure 6 | a) Relative amino acid frequency in encoded proteomes of *Homo sapiens* (UP000005640), *Staphylococcus aureus* (UP000008816) and *Pseudomonas aeruginosa* (UP000002438). Data accessed July 8th, 2021.[81] **b)** Overview of all amino acid side chains that contain potentially reactive functional groups and of the protein termini. **c)** Acylating agents targeting lysine. **d)** Salicylaldehydes like the FDA-approved drug voxelotor form stabilized imines with amines. **e)** SuFEx and SuTEx electrophiles. The site of nucleophilic attack is indicated by an orange circle.

Since the pK_a values of nucleophilic residues vary greatly depending on their microenvironment, certain residues of more abundant amino acids can also be selectively addressed. For lysine, for example, almost all residues are protonated at physiological pH but surrounding residues can result in a broad range of $pK_a \approx 5\text{-}11$.[82] Although *Michael* acceptors can also react with lysines,[83,84] not many cases of this reactivity have been reported and they harbor a significant risk of cysteine off-target reactivity. Acylating agents like activated esters (**12**)[85] and *N*-acyl-*N*-alkyl sulfonamides (**13**)[86] have been used to form stable amide bonds on lysines (Figure 6c). The reaction with salicylaldehydes (**14**) to produce stabilized imines also allows the modification of lysines[87] and, as in the case of voxelotor, the *N*-terminus (Figure 6d).[88] Notably, voxelotor is the only approved covalent drug that targets a non-catalytic residue other than cysteine.[70]

Sulfur(VI) fluoride exchange (SuFEx; **15**) chemistry has been used for the modification of non-catalytic lysines[89] but also is reactive towards other residues like non-catalytic tyrosines and catalytic serines and threonines (Figure 6e).[90] With triazoles (SuTEx; **16**) instead of fluoride as leaving group, a higher preference for tyrosines could be achieved.[91]

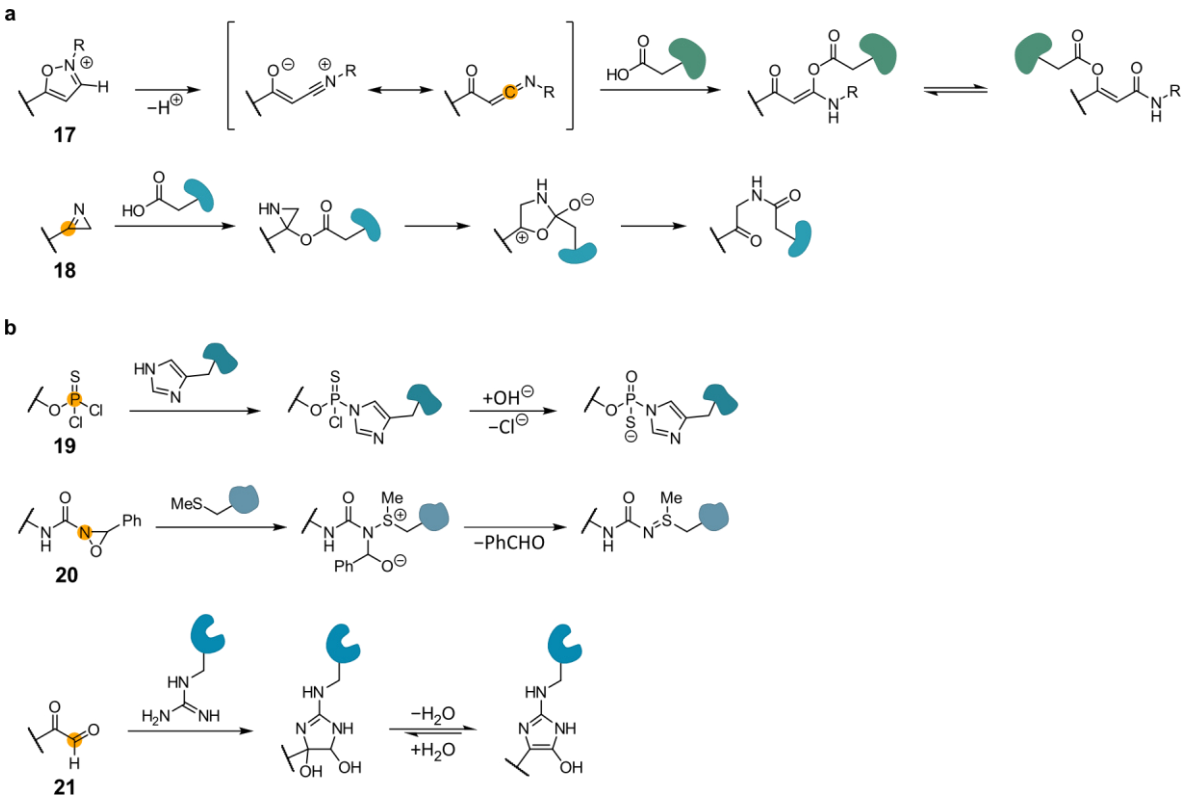


Figure 7 | a) Electrophilic motifs targeting carboxylic acid residues. **b)** Selected chemistries for bioconjugation of other amino acids.

Isoxazolium salts (**17**) react with carboxylic acid residues after base-induced fragmentation to produce enol esters (Figure 7a)[92] but have also been reported to react with histidines and cysteines.[93] Since azirines (**18**) have been reported to selectively engage aspartates and glutamates across the proteome,[94] they present a promising compound class for the development of TCIs.

While other chemistries have been developed to also engage other amino acids, they have primarily been applied in bioconjugation rather than inhibitor development. Noteworthy examples are the biomimetic thiophosphorylation of histidines (**19**),[95] redox-activated oxaziridines (**20**) for methionine modification,[96] and glyoxals (**21**) for arginine conjugation (Figure 7b).[97]

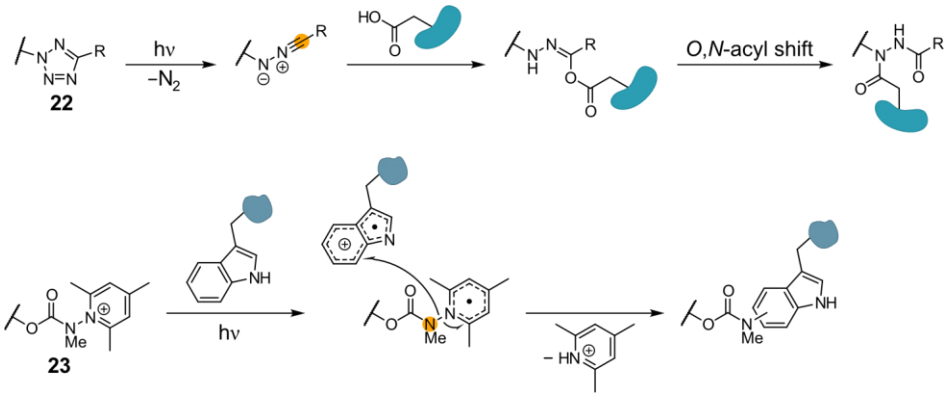


Figure 8 | Light-activated electrophiles for modification of aspartates and glutamates, and tryptophans.

Furthermore, latent electrophiles that get activated through light irradiation have been established as tool compounds. 2,5-disubstituted tetrazoles (**22**) form reactive nitrilimines that after initial reaction with aspartate or glutamate undergo an *O*,*N*-acyl shift to produce stable products (Figure 8).[98] Recently, *N*-carbamoylpyridinium salts (**23**) have been described to selectively react with tryptophans through photoinduced electron transfer.[99]

Development of TCIs

Although appending a covalent binding element to a known non-covalent inhibitor structure is an effective strategy to develop new TCIs (Figure 9), it is limited by the potentially detrimental influence of this modification on the non-covalent binding and the necessity for a suitable nucleophilic residue being present in the binding pocket.[100] Moreover, binding sites that are not tractable without covalent interaction cannot be discovered in this way.

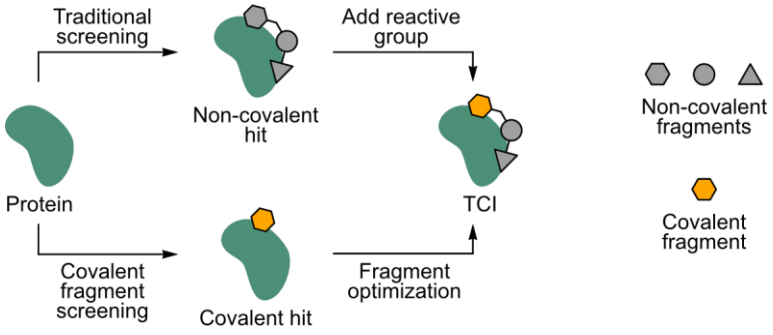


Figure 9 | Schematic approaches to TCI development.

The screening of covalent fragments offers an alternative approach. Fragments are binding elements with low molecular weight (<300 Da) for which high structural diversity can already be achieved in relatively small screening libraries compared to drug-like molecules.[101] Fragment screens provide a higher hit rate than traditional high throughput approaches and, while optimization is not trivial, fragments are considered to be better starting points for

optimization as opposed to already drug-sized hit compounds.[101,102] Fragments, furthermore, allow the identification of additional target sites beyond *e.g.* substrate binding sites.[103]

In a typical screen, a commercial fragment library is tested against one or a panel of purified protein targets.[66,75,104] As a result of their small size, interaction with the target proteins is relatively weak, which can be problematic for experimental hit identification in non-covalent libraries. Covalent fragments, however, are intrinsically suited for streamlined detection of binding through intact protein mass spectrometry (IPMS).[75,100]

While this approach allows the development of TCIs of proteins with known biological relevance, ligands of other proteins might also exert therapeutically relevant effects. In so-called inverse drug discovery, the covalent ligands are screened in living cells or cellular lysates to enable agnostic target discovery (Figure 10).[105] By employing phenotypic screening of the library and selecting the most potent compounds for target profiling, identification of covalent fragments for targets related to a particular biological question can be achieved in a highly parallel fashion.

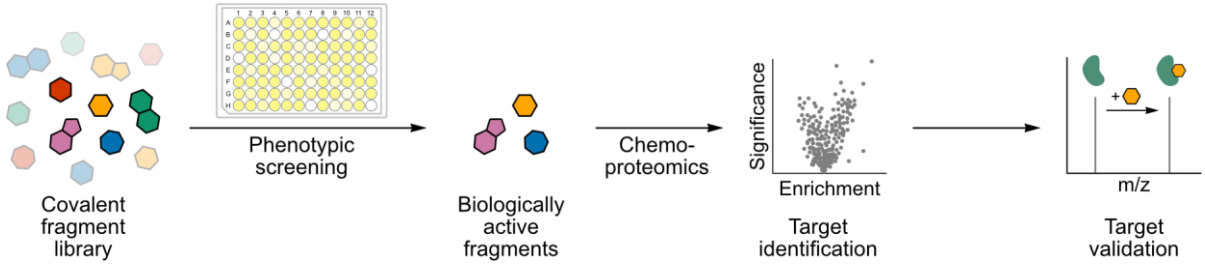


Figure 10 | Inverse drug discovery of TCIs by combination of phenotypic screening and chemoproteomics.

3. Chemoproteomics

Activity-Based Protein Profiling

Chemoproteomic methods have proven essential tools for the deconvolution of the targets of covalent ligands.[106] One of the most impactful technologies for this purpose is activity-based protein profiling (ABPP), which uses probes that combine an electrophilic motif or ligand with a reporter group. This can either be a fluorophore like rhodamine for visualization in gel electrophoresis or *e.g.* a biotin motif for affinity enrichment with immobilized (strept-)avidin (Figure 11a).[107] In the latter case, tryptic digestion and analysis of the resulting dissolved peptides by liquid chromatography coupled with mass spectrometry (LC-MS) allows the identification of target proteins.

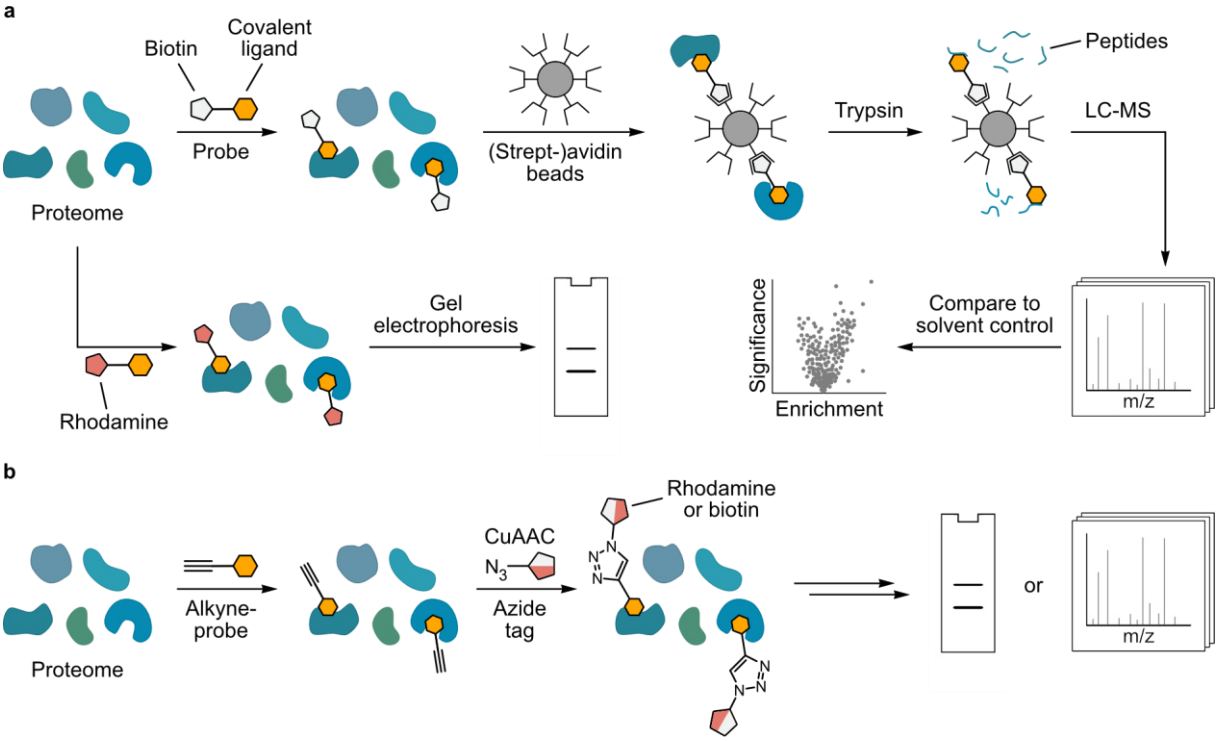


Figure 11 | a) ABPP with biotin- or rhodamine-probes. **b)** Labeling with an alkyne probe followed by CuAAC with an azide tag.

As the large and polar modification of the covalent ligand with biotin or a fluorophore can hamper the interaction with proteins and cell permeation, alkyne probes can also be used instead.[108] After the proteome is incubated with such a probe, azide tags are attached through a copper(I)-catalyzed azide-alkyne cycloaddition (CuAAC)[109] (Figure 11b). This approach furthermore shortens probe synthesis and enables the use of the same probe for different applications.[108]

The described typical ABPP approach can thus be used for the effective identification of the on- and off-targets of covalent compounds and aid drug development.[110] However, it also has

two major inherent drawbacks. The modification of the compound of interest with even just a minimal alkyne handle can perturb its binding and biological activity[111] and optimization of the modified position at which the modification can be installed, can necessitate extensive synthetic efforts. Moreover, as the probe-modified peptide remains attached to the enrichment matrix, the binding mode and the exact site of conjugation cannot be elucidated.

isoTOP-ABPP

The group of Benjamin Cravatt developed a residue-specific method, that overcomes these two limitations. For isotopic tandem orthogonal proteolysis (isoTOP)-ABPP, a biotin azide tag is used that bears an isotopically labeled peptide sequence as a linker that is cleavable by the tobacco etch virus (TEV) protease (Figure 12a).[112] As a result, the probe modified tryptic peptides can be eluted by TEV-proteolysis and analyzed by tandem mass spectrometry to identify the exact residue the modification occurred on.[113,114] The isotope labels further enable the relative quantification of a modified peptide between two differently treated samples.

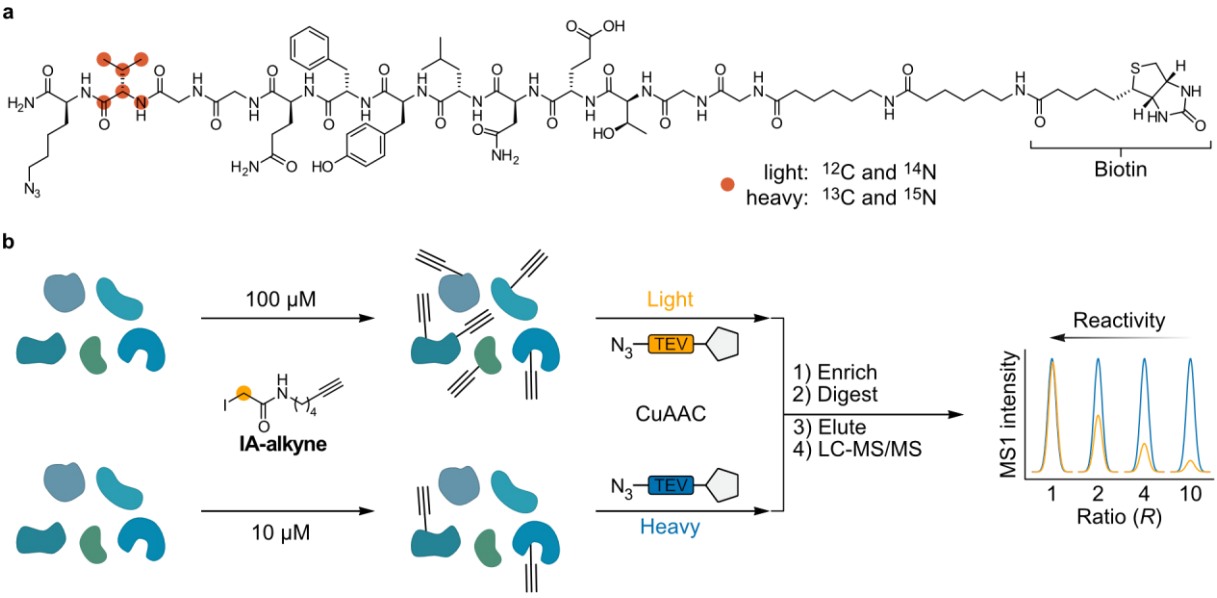


Figure 12 | a) Isotopically labeled TEV tag used in isoTOP-ABPP. **b)** Proteome-wide profiling of cysteine reactivity. The site of nucleophilic attack is indicated by an orange circle.

This method has first been used for the proteome-wide profiling of the reactivity of cysteines by treating two identical samples of the human proteome with different concentrations of the broadly reactive probe **IA-alkyne** (Figure 12b). While this probe will react with a large number of cysteines at high concentrations (light label), it will only still react quantitatively with the most reactive ones at a low concentration (heavy label). Accordingly, the extent of labeling is nearly identical for hyper-reactive cysteines, resulting in a 1:1 ratio of intensity in the heavy

and light signals (R ≈ 1), whereas less reactive cysteines will display concentration-dependent labeling (R >> 1). Notably, the identified hyper-reactive cysteines tend to carry out important catalytic and regulatory functions.[112]

This chemoproteomic platform was further adapted for competitive profiling with a covalent fragment library consisting of chloroacetamides and acrylamides.[111] For this, two identical human proteome samples were treated with either a covalent fragment (light label) or solvent control (heavy label) before treatment with a high concentration of **IA-alkyne** (Figure 13a).[111] While the extensive modification of cysteines across the proteome is unperturbed in the solvent control, some residues will already be engaged after treatment with the covalent ligand and thus blocked from reaction with the probe. The resulting difference in the proteome-wide probe labeling can again be quantified through the ratio of intensities of heavy- and light-labeled peptides and here corresponds to the target occupancy of the covalent fragment.[111] Experiments with different concentrations of the competitor can be used to directly obtain a measure for the affinity of the interaction.[115,116]

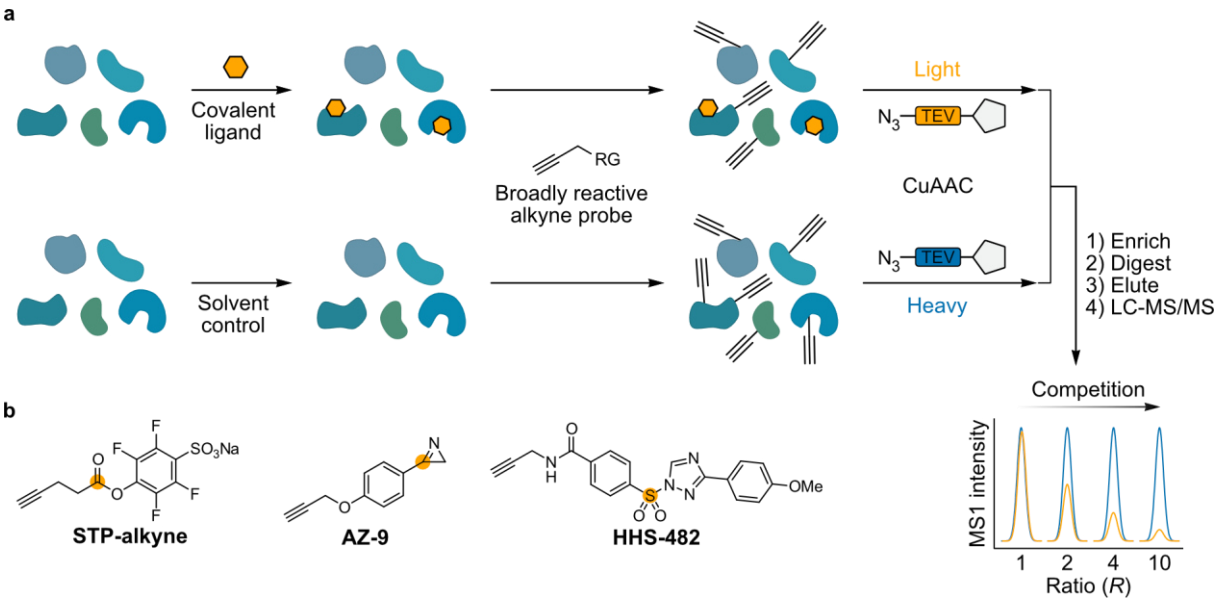


Figure 13 | a) Competitive isoTOP-ABPP workflow. **b)** Broadly reactive probes that have been used in proteome-wide ligandability studies. The site of nucleophilic attack is indicated by an orange circle.

A so-called ligandability map combines the data from these competitive experiments and visualizes the propensity of the detected cysteines to interact with members of the library and *vice versa*. Importantly, the proteome-wide pattern indicates that the interaction between covalent fragments and proteins is significantly driven by affinity of the non-covalent binding element rather than reactivity alone.[111] Besides the degree of target and fragment promiscuity, this analysis also reveals structure-activity-relationships (SARs) across the whole proteome.

The data further suggests that this approach facilitates the discovery of ligands for many targets, for which no ligands are known, including non-classic targets like transcription factors.[111]

Following the same concept, Hacker *et al.* quantified over 9000 lysines proteome-wide with **STP-alkyne** (Figure 13b) as broadly reactive probe and confirmed a small subset of all detected lysines to be hyperreactive (*cf.* Figure 12b).[85] Intriguingly, out of all the proteins that were engaged by a fragment library consisting mainly of activated esters, only ~20% were found to also carry a cysteine that was liganded in the study described above.[85] Recently, Ma *et al.* identified **AZ-9** as a probe with high selectivity for aspartates and glutamates and used it to profile the reactivity of these residues proteome-wide.[94] Even though only a preliminary competitive experiment was performed,[94] azirines show high promise for the development of carboxylic acid-directed TCIs. Using a different competitive chemoproteomic technology (stable isotope labeling by amino acids in cell culture; SILAC[117]), Brulet *et al.* investigated the ligandability of tyrosines with the broadly reactive probe **HHS-482** and fragment-based SuTEx-reagents.[118] In the two studies on lysines and tyrosines, non-classical binding sites like protein-protein interactions and non-catalytic residues were validated to be liganded by members of the fragment libraries.[85,118]. Together these results highlight the complementary nature and potential of ligand discovery targeting nucleophilic amino acids other than cysteine.

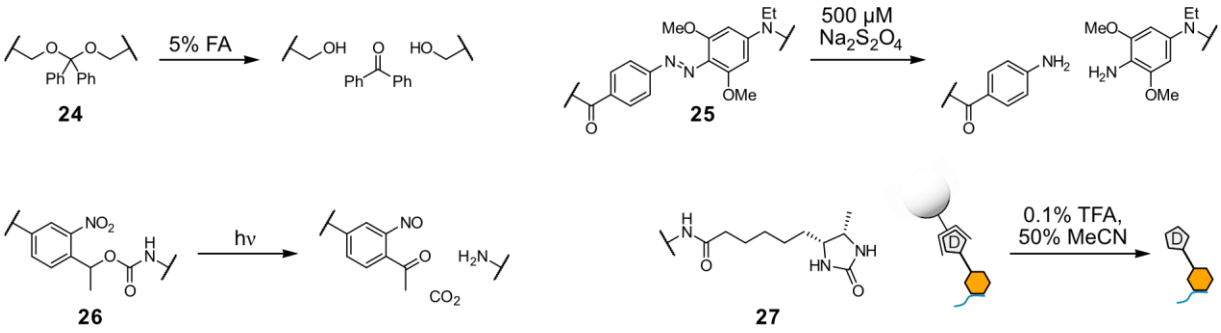


Figure 14 | Selection of cleavable linkers as well as desthiobiotin as affinity tag for elution of probe-modified peptides.

Alternative Strategies for Enrichment and Release

In addition to orthogonal proteolysis, several other methods have been reported to elute the probe-modified peptide from the enrichment matrix.[119] In general, harsh conditions and reactive intermediates are to be avoided as side reactions are undesirable and compatibility with the workflow is necessary. Linker cleavage can for example be carried out under acidic (acetal **24**)[120] or reductive conditions (azobenzene **25**)[121] or through photolysis (*o*-nitro-benzylethers **26**) (Figure 14).[122] Another approach would be to change the affinity handle to

desthiobiotin (**27**) which has a lower affinity to (strept-)avidin than biotin. The reversible binding under mildly acidic conditions facilitates elution without linker cleavage[123] and would thus be a novel feature for competitive, residue-specific proteomics.

4. Research Objectives

Establish a Chemoproteomic Platform for Development of Antibiotic TCIs

Competitive isoTOP-ABPP has been shown to streamline the identification of the targets of covalent fragments with resolution of the bound residue. As previous studies have mainly investigated the human proteome,[85,111] application of such a chemoproteomic workflow to bacterial samples is a promising approach to discover new antibiotic TCIs as well as target proteins. For this purpose, an alternative isotopically labeled azide tag utilizing desthiobiotin for enrichment will be developed (Figure 15). Since no (photo-)chemically or proteolytically cleavable motifs are required, this will facilitate a simple and quick tag synthesis that is compatible with a wide variety of chemistries and conditions and thus maximize the versatility of the method. After successful synthesis and favorable benchmarking of these isotopically labeled desthiobiotin azide (isoDTB) tags, the reactivity and ligandability of cysteines in the bacterial proteome will be investigated. For this, a commercial covalent fragment library will be tested for antibiotic activity and the most potent compounds will be selected for target identification (*cf.* Figure 10). Finally, promising hits will be validated in *in vitro* studies like IPMS.

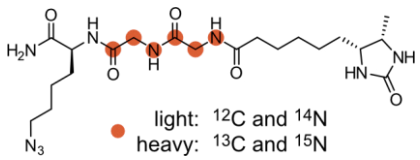


Figure 15 | Structure of isoDTB tags.

Extensive Profiling of Nucleophilic Residues in the Bacterial Proteome

After establishing isoDTB-ABPP for the profiling of cysteines, efforts will be directed to profile other nucleophilic amino acid residues in order to identify additional sites that can be addressed by TCIs.

Aspartates and glutamates together make up ~12% of the bacterial proteome[81] (Figure 6a) and their carboxylic acid side chains offer the possibility of selective modification in the presence of other nucleophiles. The suitability of different photoactivatable 2,5-disubstituted tetrazoles[98] (Figure 8) as broadly reactive alkyne probes will be evaluated. While only few covalent ligands targeting carboxylic acid residues have been reported, competitive experiments with isoxazolium salts (Figure 7a) will be carried out as a proof of concept study.[92] Due to the property of the tetrazole probes to be non-reactive and thus probably non-toxic before the irradiation, they can also be used to study the native environment of living

cells instead of cellular lysates.[124] Accordingly, the monitoring of aspartates and glutamates in living bacterial cells will be evaluated as well.

Broadly reactive probes that are selective towards a particular side chain functionality are crucial for the exploration of covalently ligandable space with competitive, residue-specific proteomics and the development of TCIs. While many electrophiles have been developed to address the various nucleophilic amino acid side chains,[71] their reactivity and selectivity in the complex setting of cellular lysates or live cells were rarely investigated and different bodies of work are not directly comparable with each other. In collaboration with other groups, this issue will be addressed by the synthesis of a large variety of different electrophilic probes and their subsequent evaluation in a unified workflow. Furthermore, the bioinformatic platform FragPipe[125,126] will be tailored for the analysis of the resulting masses of modification, their site selectivity and their quantification.

5. References

[1] R. Laxminarayan *et al.*, Access to effective antimicrobials, *Lancet* **2016**, *387*, 168–175.

[2] Interagency Coordination Group on Antimicrobial Resistance, No Time to Wait: Securing the Future from Drug-Resistant Infections **2019**.

[3] K. E. Fleming-Dutra *et al.*, Prevalence of Inappropriate Antibiotic Prescriptions Among US Ambulatory Care Visits, 2010-2011, *JAMA* **2016**, *315*, 1864–1873.

[4] H. Zhao *et al.*, Appropriateness of antibiotic prescriptions in ambulatory care in China, *Lancet Infect. Dis.* **2021**, *21*, 847–857.

[5] T. P. van Boeckel *et al.*, Global trends in antimicrobial resistance in animals in low- and middle-income countries, *Science* **2019**, *365*.

[6] B. M. Marshall, S. B. Levy, Food animals and antimicrobials, *Clin. Microbiol. Rev.* **2011**, *24*, 718–733.

[7] World Health Organization, *2020 Antibacterial Agents in Clinical and Preclinical Development:* an overview and analysis, **2021**.

[8] C. Årdal *et al.*, Antibiotic development - economic, regulatory and societal challenges, *Nat. Rev. Microbiol.* **2020**, *18*, 267–274.

[9] U. Theuretzbacher *et al.*, Analysis of the clinical antibacterial and antituberculosis pipeline, *Lancet Infect. Dis.* **2019**, *19*, e40-e50.

[10] U. Theuretzbacher, K. Outterson, A. Engel, A. Karlén, The global preclinical antibacterial pipeline, *Nat. Rev. Microbiol.* **2020**, *18*, 275–285.

[11] Eastern Research Group, Analytical Framework for Examining the Value of Antibacterial Products **2014**.

[12] S. Hernando-Amado, T. M. Coque, F. Baquero, J. L. Martínez, Defining and combating antibiotic resistance from One Health and Global Health perspectives, *Nat. Microbiol.* **2019**, *4*, 1432–1442.

[13] B. Oronsky *et al.*, A Review of Persistent Post-COVID Syndrome (PPCS), *Clin. Rev. Allergy Immunol.* **2021**.

[14] A. Abbott, COVID's mental-health toll, *Nature* **2021**, *590*, 194–195.

[15] D. E. Morris, D. W. Cleary, S. C. Clarke, Secondary Bacterial Infections Associated with Influenza Pandemics, *Front. Microbiol.* **2017**, *8*, 1041.

[16] M. A. B. Lucien *et al.*, Antibiotics and antimicrobial resistance in the COVID-19 era, *Int. J. Infect. Dis.* **2021**, *104*, 250–254.

[17] C. D. Russell *et al.*, Co-infections, secondary infections, and antimicrobial use in patients hospitalised with COVID-19 during the first pandemic wave from the ISARIC WHO CCP-UK study, *Lancet Infect. Dis.* **2021**, *10*, 394.

[18] Review on Antimicrobial Resistance, *Antimicrobial Resistance:* Tackling a crisis for the health and wealth of nations, **2014**.

[19] Review on Antimicrobial Resistance, *Tackling Drug-Resistant Infections Globally:* Final Report and Recommendations, **2016**.

[20] World Bank Group, *Drug-Resistant Infections:* A Threat to Our Economic Future, **2017**.

[21] R. C. Moellering, NDM-1--a cause for worldwide concern, *N. Engl. J. Med.* **2010**, *363*, 2377–2379.

[22] M.-E. Böhm, M. Razavi, N. P. Marathe, C.-F. Flach, D. G. J. Larsson, Discovery of a novel integron-borne aminoglycoside resistance gene present in clinical pathogens by screening environmental bacterial communities, *Microbiome* **2020**, *8*, 41.

[23] J. Cama *et al.*, To Push or To Pull? In a Post-COVID World, Supporting and Incentivizing Antimicrobial Drug Development Must Become a Governmental Priority, *ACS Infect. Dis.* **2021**.

[24] G. Alvarez-Uria, S. Gandra, R. Laxminarayan, Poverty and prevalence of antimicrobial resistance in invasive isolates, *Int. J. Infect. Dis.* **2016**, *52*, 59–61.

[25] R. Laxminarayan, D. L. Heymann, Challenges of drug resistance in the developing world, *BMJ* **2012**, *344*, e1567.

[26] Z. Dhesi, V. I. Enne, J. O'Grady, V. Gant, D. M. Livermore, Rapid and Point-of-Care Testing in Respiratory Tract Infections, *ACS Pharmacol. Transl. Sci.* **2020**, *3*, 401–417.

[27] A. L. Hicks *et al.*, Targeted surveillance strategies for efficient detection of novel antibiotic resistance variants, *eLife* **2020**, 9:e56367.

[28] R. W. Peeling, D. Mabey, Point-of-care tests for diagnosing infections in the developing world, *Clin. Microbiol. Infect.* **2010**, *16*, 1062–1069.

[29] M. Baym, L. K. Stone, R. Kishony, Multidrug evolutionary strategies to reverse antibiotic resistance, *Science* **2016**, *351*, aad3292.

[30] F. Peláez, The historical delivery of antibiotics from microbial natural products--can history repeat?, *Biochem. Pharmacol.* **2006**, *71*, 981–990.

[31] L. L. Ling *et al.*, A new antibiotic kills pathogens without detectable resistance, *Nature* **2015**, *517*, 455–459.

[32] M. F. Richter *et al.*, Predictive compound accumulation rules yield a broad-spectrum antibiotic, *Nature* **2017**, *545*, 299–304.

[33] H. I. Zgurskaya, C. A. Löpez, S. Gnanakaran, Permeability Barrier of Gram-Negative Cell Envelopes and Approaches To Bypass It, *ACS Infect. Dis.* **2015**, *1*, 512–522.

[34] X.-Z. Li, H. Nikaido, Efflux-mediated drug resistance in bacteria, *Drugs* **2004**, *64*, 159–204.

[35] E. N. Parker *et al.*, Implementation of permeation rules leads to a FabI inhibitor with activity against Gram-negative pathogens, *Nat. Microbiol.* **2020**, *5*, 67–75.

[36] S. J. Perlmutter *et al.*, Compound Uptake into E. coli Can Be Facilitated by N-Alkyl Guanidiniums and Pyridiniums, *ACS Infect. Dis.* **2021**, *7*, 162–173.

[37] M. Lakemeyer, W. Zhao, F. A. Mandl, P. Hammann, S. A. Sieber, Thinking Outside the Box-Novel Antibacterials To Tackle the Resistance Crisis, *Angew. Chem. Int. Ed.* **2018**, *57*, 14440–14475.

[38] R. Zhang, Y. Lin, DEG 5.0, a database of essential genes in both prokaryotes and eukaryotes, *Nucleic Acids Res.* **2009**, *37*, D455-8.

[39] K. Lewis, The Science of Antibiotic Discovery, *Cell* **2020**, *181*, 29–45.

[40] S. W. Dickey, G. Y. C. Cheung, M. Otto, Different drugs for bad bugs: antivirulence strategies in the age of antibiotic resistance, *Nat. Rev. Drug Discov.* **2017**, *16*, 457–471.

[41] F. L. Gordillo Altamirano, J. J. Barr, Phage Therapy in the Postantibiotic Era, *Clin. Microbiol. Rev.* **2019**, *32*.

[42] World Health Organization, *World Health Organization Model List of Essential Medicines:* 21st List, **2019**.

[43] A. L. Demain, R. P. Elander, The beta-lactam antibiotics, *Antonie van Leeuwenhoek* **1999**, *75*, 5–19.

[44] M. E. Falagas, E. K. Vouloumanou, G. Samonis, K. Z. Vardakas, Fosfomycin, *Clin. Microbiol. Rev.* **2016**, *29*, 321–347.

[45] J. Singh, R. C. Petter, T. A. Baillie, A. Whitty, The resurgence of covalent drugs, *Nat. Rev. Drug Discov.* **2011**, *10*, 307–317.

[46] S. de Cesco, J. Kurian, C. Dufresne, A. K. Mittermaier, N. Moitessier, Covalent inhibitors design and discovery, *Eur. J. Med. Chem.* **2017**, *138*, 96–114.

[47] H. G. Bull *et al.*, Mechanism-Based Inhibition of Human Steroid 5α-Reductase by Finasteride: Enzyme-Catalyzed Formation of NADP−Dihydrofinasteride, a Potent Bisubstrate Analog Inhibitor, *J. Am. Chem. Soc.* **1996**, *118*, 2359–2365.

[48] J. F. Steiner, Clinical pharmacokinetics and pharmacodynamics of finasteride, *Clin. Pharmacokinet.* **1996**, *30*, 16–27.

[49] A. J. Claxton, J. Cramer, C. Pierce, A systematic review of the associations between dose regimens and medication compliance, *Clin. Ther.* **2001**, *23*, 1296–1310.

[50] R. A. Copeland, The drug-target residence time model: a 10-year retrospective, *Nat. Rev. Drug Discov.* **2016**, *15*, 87–95.

[51] A. Abdeldayem, Y. S. Raouf, S. N. Constantinescu, R. Moriggl, P. T. Gunning, Advances in covalent kinase inhibitors, *Chem. Soc. Rev.* **2020**, *49*, 2617–2687.

[52] D. J. Tipper, J. L. Strominger, Mechanism of action of penicillins: a proposal based on their structural similarity to acyl-D-alanyl-D-alanine, *Proc. Natl. Acad. Sci. USA* **1965**, *54*, 1133–1141.

[53] S. Ray, A. S. Murkin, New Electrophiles and Strategies for Mechanism-Based and Targeted Covalent Inhibitor Design, *Biochemistry* **2019**, *58*, 5234–5244.

[54] J. Way, Covalent modification as a strategy to block protein–protein interactions with small-molecule drugs, *Curr. Opin. Chem. Biol.* **2000**, *4*, 40–46.

[55] S.-S. Cheng, G.-J. Yang, W. Wang, C.-H. Leung, D.-L. Ma, The design and development of covalent protein-protein interaction inhibitors for cancer treatment, *J. Hematol. Oncol.* **2020**, *13*, 26.

[56] J. Engel, J. Lategahn, D. Rauh, Hope and Disappointment: Covalent Inhibitors to Overcome Drug Resistance in Non-Small Cell Lung Cancer, *ACS Med. Chem. Lett.* **2016**, *7*, 2–5.

[57] K. S. Thress *et al.*, Acquired EGFR C797S mutation mediates resistance to AZD9291 in non-small cell lung cancer harboring EGFR T790M, *Nat. Med.* **2015**, *21*, 560–562.

[58] Y. Chen, Y. Jia, W. Song, L. Zhang, Therapeutic Potential of Nitrogen Mustard Based Hybrid Molecules, *Front. Pharmacol.* **2018**, *9*, 1453.

[59] F. Sutanto, M. Konstantinidou, A. Dömling, Covalent inhibitors: a rational approach to drug discovery, *RSC Med. Chem.* **2020**, *11*, 876–884.

[60] D. Li *et al.*, BIBW2992, an irreversible EGFR/HER2 inhibitor highly effective in preclinical lung cancer models, *Oncogene* **2008**, *27*, 4702–4711.

[61] B. A. Lanman *et al.*, Discovery of a Covalent Inhibitor of KRASG12C (AMG 510) for the Treatment of Solid Tumors, *J. Med. Chem.* **2020**, *63*, 52–65.

[62] J. Lei, Y. Zhou, D. Xie, Y. Zhang, Mechanistic insights into a classic wonder drug--aspirin, *J. Am. Chem. Soc.* **2015**, *137*, 70–73.

[63] J. L. Andersen *et al.*, Dimethyl fumarate is an allosteric covalent inhibitor of the p90 ribosomal S6 kinases, *Nat. Commun.* **2018**, *9*, 4344.

[64] *FDA Approves First Targeted Therapy for Lung Cancer Mutation Previously Considered Resistant to Drug Therapy*, U.S. Food and Drug Administration, **2021**, can be found under www.fda.gov/news-events/press-announcements/fda-approves-first-targeted-therapy-lung-cancer-mutation-previously-considered-resistant-drug (accessed Jul 8, 2021).

[65] A. D. Cox, S. W. Fesik, A. C. Kimmelman, J. Luo, C. J. Der, Drugging the undruggable RAS: Mission possible?, *Nat. Rev. Drug Discov.* **2014**, *13*, 828–851.

[66] J. M. Ostrem, U. Peters, M. L. Sos, J. A. Wells, K. M. Shokat, K-Ras(G12C) inhibitors allosterically control GTP affinity and effector interactions, *Nature* **2013**, *503*, 548–551.

[67] A. Tuley, W. Fast, The Taxonomy of Covalent Inhibitors, *Biochemistry* **2018**, *57*, 3326–3337.

[68] Y.-H. Wang, F. Zhang, H. Diao, R. Wu, Covalent Inhibition Mechanism of Antidiabetic Drugs—Vildagliptin vs Saxagliptin, *ACS Catal.* **2019**, *9*, 2292–2302.

[69] D. T. King, A. M. King, S. M. Lal, G. D. Wright, N. C. J. Strynadka, Molecular Mechanism of Avibactam-Mediated β-Lactamase Inhibition, *ACS Infect. Dis.* **2015**, *1*, 175–184.

[70] E. de Vita, 10 years into the resurgence of covalent drugs, *Future Med. Chem.* **2021**, *13*, 193–210.

[71] M. Gehringer, S. A. Laufer, Emerging and Re-Emerging Warheads for Targeted Covalent Inhibitors, *J. Med. Chem.* **2019**, *62*, 5673–5724.

[72] A. Birkholz *et al.*, Systematic Study of the Glutathione Reactivity of N-Phenylacrylamides: 2. Effects of Acrylamide Substitution, *J. Med. Chem.* **2020**, *63*, 11602–11614.

[73] I. M. Serafimova *et al.*, Reversible targeting of noncatalytic cysteines with chemically tuned electrophiles, *Nat. Chem. Biol.* **2012**, *8*, 471–476.

[74] J. K. Eaton, L. Furst, L. L. Cai, V. S. Viswanathan, S. L. Schreiber, Structure-activity relationships of GPX4 inhibitor warheads, *Bioorg. Med. Chem. Lett.* **2020**, *30*, 127538.

[75] E. Resnick *et al.*, Rapid Covalent-Probe Discovery by Electrophile-Fragment Screening, *J. Am. Chem. Soc.* **2019**, *141*, 8951–8968.

[76] C. Dubiella *et al.*, Sulfopin is a covalent inhibitor of Pin1 that blocks Myc-driven tumors in vivo, *Nat. Chem. Biol.* **2021**.

[77] D. Allimuthu, D. J. Adams, 2-Chloropropionamide As a Low-Reactivity Electrophile for Irreversible Small-Molecule Probe Identification, *ACS Chem. Biol.* **2017**, *12*, 2124–2131.

[78] N. Shindo *et al.*, Selective and reversible modification of kinase cysteines with chlorofluoroacetamides, *Nat. Chem. Biol.* **2019**, *15*, 250–258.

[79] R. Gianatassio *et al.*, Strain-release amination, *Science* **2016**, *351*, 241–246.

[80] A. M. Embaby, S. Schoffelen, C. Kofoed, M. Meldal, F. Diness, Rational Tuning of Fluorobenzene Probes for Cysteine-Selective Protein Modification, *Angew. Chem. Int. Ed.* **2018**, *57*, 8022–8026.

[81] UniProt: a hub for protein information, *Nucleic Acids Res.* **2015**, *43*, D204-12.

[82] D. G. Isom, C. A. Castañeda, B. R. Cannon, B. García-Moreno, Large shifts in pKa values of lysine residues buried inside a protein, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 5260–5265.

[83] J. Pettinger *et al.*, An Irreversible Inhibitor of HSP72 that Unexpectedly Targets Lysine-56, *Angew. Chem. Int. Ed.* **2017**, *56*, 3536–3540.

[84] E. Anscombe *et al.*, Identification and Characterization of an Irreversible Inhibitor of CDK2, *Chem. Biol.* **2015**, *22*, 1159–1164.

[85] S. M. Hacker *et al.*, Global profiling of lysine reactivity and ligandability in the human proteome, *Nat. Chem.* **2017**, *9*, 1181–1190.

[86] T. Tamura *et al.*, Rapid labelling and covalent inhibition of intracellular native proteins using ligand-directed N-acyl-N-alkyl sulfonamide, *Nat. Commun.* **2018**, *9*, 1870.

[87] M. Sanches *et al.*, Structure and mechanism of action of the hydroxy-aryl-aldehyde class of IRE1 endoribonuclease inhibitors, *Nat. Commun.* **2014**, *5*, 4202.

[88] B. Metcalf *et al.*, Discovery of GBT440, an Orally Bioavailable R-State Stabilizer of Sickle Cell Hemoglobin, *ACS Med. Chem. Lett.* **2017**, *8*, 321–326.

[89] N. P. Grimster *et al.*, Aromatic sulfonyl fluorides covalently kinetically stabilize transthyretin to prevent amyloidogenesis while affording a fluorescent conjugate, *J. Am. Chem. Soc.* **2013**, *135*, 5656–5668.

[90] A. Narayanan, L. H. Jones, Sulfonyl fluorides as privileged warheads in chemical biology, *Chem. Sci.* **2015**, *6*, 2650–2659.

[91] A. L. Borne, J. W. Brulet, K. Yuan, K.-L. Hsu, Development and biological applications of sulfur-triazole exchange (SuTEx) chemistry, *RSC Chem. Biol.* **2021**, *2*, 322–337.

[92] P. Martín-Gago *et al.*, Covalent Protein Labeling at Glutamic Acids, *Cell Chem. Biol.* **2017**, *24*, 589-597.e5.

[93] P. Bustos *et al.*, Woodward's reagent K reacts with histidine and cysteine residues in Escherichia coli and Saccharomyces cerevisiae phosphoenolpyruvate carboxykinases, *J. Protein. Chem.* **1996**, *15*, 467–472.

[94] N. Ma *et al.*, 2H-Azirine-Based Reagents for Chemoselective Bioconjugation at Carboxyl Residues Inside Live Cells, *J. Am. Chem. Soc.* **2020**, *142*, 6051–6059.

[95] S. Jia, D. He, C. J. Chang, Bioinspired Thiophosphorodichloridate Reagents for Chemoselective Histidine Bioconjugation, *J. Am. Chem. Soc.* **2019**, *141*, 7294–7301.

[96] S. Lin *et al.*, Redox-based reagents for chemoselective methionine bioconjugation, *Science* **2016**, *355*, 597–602.

[97] I. Dovgan *et al.*, Arginine-selective bioconjugation with 4-azidophenyl glyoxal: application to the single and dual functionalisation of native antibodies, *Org. Biomol. Chem.* **2018**, *16*, 1305–1311.

[98] S. Zhao *et al.*, Photo-induced coupling reactions of tetrazoles with carboxylic acids in aqueous solution: application in protein labelling, *Chem. Commun.* **2016**, *52*, 4702–4705.

[99] S. J. Tower, W. J. Hetcher, T. E. Myers, N. J. Kuehl, M. T. Taylor, Selective Modification of Tryptophan Residues in Peptides and Proteins Using a Biomimetic Electron Transfer Process, *J. Am. Chem. Soc.* **2020**, *142*, 9112–9118.

[100] W. Lu *et al.*, Fragment-based covalent ligand discovery, *RSC Chem. Biol.* **2021**, *2*, 354–367.

[101] C. W. Murray, D. C. Rees, The rise of fragment-based drug discovery, *Nat. Chem.* **2009**, *1*, 187–192.

[102] S. D. Bembenek, B. A. Tounge, C. H. Reynolds, Ligand efficiency and fragment-based drug discovery, *Drug Discov. Today* **2009**, *14*, 278–283.

[103] R. F. Ludlow, M. L. Verdonk, H. K. Saini, I. J. Tickle, H. Jhoti, Detection of secondary binding sites in proteins using fragment screening, *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 15910–15915.

[104] L. Petri *et al.*, Assessment of Tractable Cysteines for Covalent Targeting by Screening Covalent Fragments, *ChemBioChem* **2021**, *22*, 743–753.

[105] D. E. Mortenson *et al.*, "Inverse Drug Discovery" Strategy To Identify Proteins That Are Targeted by Latent Electrophiles As Exemplified by Aryl Fluorosulfates, *J. Am. Chem. Soc.* **2018**, *140*, 200–210.

[106] L. P. Conway, W. Li, C. G. Parker, Chemoproteomic-enabled phenotypic screening, *Cell Chem. Biol.* **2021**, *28*, 371–393.

[107] Y. Liu, M. P. Patricelli, B. F. Cravatt, Activity-based protein profiling: the serine hydrolases, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 14694–14699.

[108] J. Martell, E. Weerapana, Applications of copper-catalyzed click chemistry in activity-based protein profiling, *Molecules* **2014**, *19*, 1378–1393.

[109] V. V. Rostovtsev, L. G. Green, V. V. Fokin, K. B. Sharpless, A Stepwise Huisgen Cycloaddition Process: Copper(I)-Catalyzed Regioselective "Ligation" of Azides and Terminal Alkynes, *Angew. Chem. Int. Ed.* **2002**, *41*, 2596–2599.

[110] C. M. Pichler, J. Krysiak, R. Breinbauer, Target identification of covalently binding drugs by activity-based protein profiling (ABPP), *Bioorg. Med. Chem.* **2016**, *24*, 3291–3303.

[111] K. M. Backus *et al.*, Proteome-wide covalent ligand discovery in native biological systems, *Nature* **2016**, *534*, 570–574.

[112] E. Weerapana *et al.*, Quantitative reactivity profiling predicts functional cysteines in proteomes, *Nature* **2010**, *468*, 790–795.

[113] E. Weerapana, G. M. Simon, B. F. Cravatt, Disparate proteome reactivity profiles of carbon electrophiles, *Nat. Chem. Biol.* **2008**, *4*, 405–407.

[114] A. E. Speers, B. F. Cravatt, A tandem orthogonal proteolysis strategy for high-content chemical proteomics, *J. Am. Chem. Soc.* **2005**, *127*, 10018–10019.

[115] G. Drewes, S. Knapp, Chemoproteomics and Chemical Probes for Target Discovery, *Trends Biotechnol.* **2018**, *36*, 1275–1286.

[116] K. M. Lee, P. Le, S. A. Sieber, S. M. Hacker, Degrasyn exhibits antibiotic activity against multi-resistant Staphylococcus aureus by modifying several essential cysteines, *Chem. Commun.* **2020**, *56*, 2929–2932.

[117] M. Mann, Functional and quantitative proteomics using SILAC, *Nat. Rev. Mol. Cell. Biol.* **2006**, *7*, 952–958.

[118] J. W. Brulet, A. L. Borne, K. Yuan, A. H. Libby, K.-L. Hsu, Liganding Functional Tyrosine Sites on Proteins Using Sulfur-Triazole Exchange Chemistry, *J. Am. Chem. Soc.* **2020**, *142*, 8270–8280.

[119] H. A. Beard, D. Korovesis, S. Chen, S. H. L. Verhelst, Cleavable linkers and their application in MS-based target identification, *Mol. Omics* **2021**, *17*, 197–209.

[120] J. Wang *et al.*, Mapping sites of aspirin-induced acetylations in live cells by quantitative acid-cleavable activity-based protein profiling (QA-ABPP), *Sci. Rep.* **2015**, *5*, 7896.

[121] T. Eom, A. Khan, Hypersensitive azobenzenes: facile synthesis of clickable and cleavable azo linkers with tunable and high reducibility, *Org. Biomol. Chem.* **2020**, *18*, 420–424.

[122] J. Li *et al.*, An Isotope-Coded Photocleavable Probe for Quantitative Profiling of Protein O-GlcNAcylation, *ACS Chem. Biol.* **2019**, *14*, 4–10.

[123] M. P. Patricelli *et al.*, In situ kinase profiling reveals functionally relevant properties of native kinases, *Chem. Biol.* **2011**, *18*, 699–710.

[124] M. Abo, E. Weerapana, A Caged Electrophilic Probe for Global Analysis of Cysteine Reactivity in Living Cells, *J. Am. Chem. Soc.* **2015**, *137*, 7087–7090.

[125] A. T. Kong, F. V. Leprevost, D. M. Avtonomov, D. Mellacheruvu, A. I. Nesvizhskii, MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics, *Nat. Methods* **2017**, *14*, 513–520.

[126] F. Yu *et al.*, Identification of modified peptides using localization-aware open search, *Nat. Commun.* **2020**, *11*, 4065.

II. Research

1. Isotopically Labeled Desthiobiotin Azide (isoDTB) Tags Enable Global Profiling of the Bacterial Cysteinome

Published in *Angewandte Chemie International Edition* **2020**, *59*, 2829-2836.

By Patrick R. A. Zanon, Lisa Lewald, Stephan M. Hacker

https://onlinelibrary.wiley.com/doi/abs/10.1002/anie.201912075

Synopsis

Most chemoproteomic studies employ biotin as an affinity handle for enrichment, which almost irreversibly binds to (strept-)avidin. This, however, necessitates the use of linkers that can be cleaved through *e.g.* proteolysis, irradiation, or reduction for the elution from the solid phase. Rather than relying on cleavable linkers, we envisioned the use of isotopically labeled tags bearing a desthiobiotin moiety. Its affinity for (strept-)avidin is still strong enough for highly specific enrichment but can be significantly decreased under relatively mild acidic conditions. Without a cleavable motif, our isoDTB tags are structurally very simple and can be synthesized by solid-phase peptide synthesis in four coupling steps with an overall yield of around 70%.

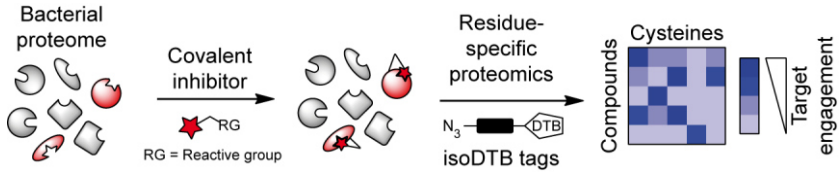
First, we established that the isoDTB tags allow for the quantification of cysteine residues in the lysate of *Staphylococcus aureus* with **IA-alkyne** as a broadly reactive probe. In direct comparison, we quantified 27% more sites than with the commonly used TEV tags of the original isoTOP-ABPP platform. By using the proteases chymotrypsin or AspN instead of trypsin, we were also able to quantify additional sites for the isoDTB tags, whereas the TEV tags are not compatible with these alternative proteases. We furthermore demonstrated that our platform effectively facilitates the profiling of cysteines in the lysates of different Gram-positive and Gram-negative bacteria, as well as the human cancer cell line MDA-MB-231.

For the human proteome, the reactivity of cysteines had been shown to correlate with their functional relevance. To study cysteine reactivity in bacteria, we treated two identical samples of *S. aureus* lysate with either 10 µM or 100 µM of **IA-alkyne**, followed by CuAAC with the light and heavy isoDTB tags, respectively. Cysteines of high reactivity are still quantitatively labeled at the lower concentration, whereas for less reactive cysteines, labeling will be significantly lower than in the sample treated with high probe concentration. We confirmed that cysteines with increased reactivity are more likely to be found in functional sites but are less common in essential proteins. A plausible explanation for the latter observation is that highly reactive residues are evolutionarily disfavored in essential proteins as they might react too unspecifically and in this way cause detrimental effects.

Next, we screened a commercial, cysteine-directed library of over 200 covalent fragments for antibacterial activity. We selected 24 compounds based on antibacterial activity and structural diversity and tested them in a competitive isoDTB-ABPP workflow. We found that the compounds have varying degrees of target promiscuity, which does not correlate with their antibacterial activity. After excluding the excessively reactive compounds, we combined the data sets to a global overview of the ligandability of 1756 cysteines in 905 proteins, which includes 59% of all cysteines in essential proteins. Importantly, among the >200 ligandable cysteines we detected many residues of low and medium reactivity, suggesting that the protein-ligand interaction is depending on non-covalent interactions rather than only on the reactivity of the fragment.

To demonstrate that the ligandability map serves as a resource to identify molecular leads for bacterial targets, we validated the interaction of compound **EN106** and C111 of the putative HMG-CoA synthase, which catalyzes the second step of the essential mevalonate pathway. In a gel-based experiment, we showed concentration-dependent competition of **EN106** with **IA-alkyne** as a probe and confirmed the expected mass of modification of the wildtype protein through IPMS. Importantly, a respective cysteine-to-alanine mutant did not show any modification, supporting C111 as the site of modification. Finally, the addition of **EN106** inhibited the enzymatic activity of HMG-CoA synthase almost completely.

Overall, this work describes the development of isoDTB tags which improve the coverage of the bacterial proteome compared to traditionally used TEV tags and obviate the need for an additional proteolytic digest, shortening the workflow. The isoDTB platform was used to assess the reactivity and, in combination with phenotypic screening, the ligandability of cysteines in *S. aureus*. The resulting data is useful for the development of antibiotics with new modes of action as shown by **EN106** targeting a so far clinically unexplored metabolic pathway.



Author Contributions

Patrick R. A. Zanon synthesized **IA-alkyne**, cloned and expressed enzymes, and performed target validation studies. Patrick R. A. Zanon and Stephan M. Hacker designed the commercial library and analyzed bioactivity and chemoproteomic data. Lisa Lewald synthesized isoDTB tags. Stephan M. Hacker tested for bioactivity and performed chemoproteomics experiments. Stephan M. Hacker wrote the manuscript with contributions from all authors.

Protein Modifications

International Edition: DOI: 10.1002/anie.201912075
German Edition: DOI: 10.1002/ange.201912075

Isotopically Labeled Desthiobiotin Azide (isoDTB) Tags Enable Global Profiling of the Bacterial Cysteinome**

*Patrick R. A. Zanon, Lisa Lewald, and Stephan M. Hacker**

Abstract: *Rapid development of bacterial resistance has led to an urgent need to find new druggable targets for antibiotics. In this context, residue-specific chemoproteomic approaches enable proteome-wide identification of binding sites for covalent inhibitors. Described here are easily synthesized isotopically labeled desthiobiotin azide (isoDTB) tags that shortened the chemoproteomic workflow and allowed an increased coverage of cysteines in bacterial systems. They were used to quantify 59% of all cysteines in essential proteins in Staphylococcus aureus and enabled the discovery of 88 cysteines that showed high reactivity, which correlates with functional importance. Furthermore, 268 cysteines that are engaged by covalent ligands were identified. Inhibition of HMG-CoA synthase was verified and will allow addressing the bacterial mevalonate pathway through a new target. Overall, a broad map of the bacterial cysteinome was obtained, which will facilitate the development of antibiotics with novel modes-of-action.*

Introduction

Infections with multidrug-resistant bacteria like methicillin-resistant *Staphylococcus aureus* (MRSA) are emerging as major threats to human health.[1] Nevertheless, very few novel classes of antibiotics have been introduced to clinics over the last decades.[1] Furthermore, almost all approved antibiotics exclusively interfere with a very limited set of bacterial targets involved in protein, nucleic acid, and cell wall biosynthesis.[1] Therefore, developing innovative methods to discover novel druggable targets for antibiotics is a pivotal task to guarantee efficient treatment of bacterial infections in the future.

Chemoproteomic approaches are extremely powerful for understanding which proteins are able to bind small molecules as ligands[2] and are particularly straightforward for covalently reactive molecules.[2a,c,d] Strikingly, covalent inhibitors have seen a resurgence of interest for the development of novel drugs as they can increase compound selectivity, reduce resistance formation, and target shallow protein pockets.[3] This interest has led to the recent clinical approval of several covalent kinase inhibitors.[4] Especially in the field of antibiotics, covalent inhibitors are prevalent as exemplified by β-lactams,[3] fosfomycin,[5] showdomycin,[6] and optimized arylomycins.[7]

Recently, methods have emerged to globally identify the exact interaction site of covalent inhibitors in a competitive fashion.[2a,b,8] In this way, many pockets that can bind covalent ligands are identified in parallel using a small library of covalently reactive molecules. This technology is especially well established for profiling cysteine residues using methods based on the isoTOP-ABPP (isotopic tandem orthogonal proteolysis activity-based protein profiling) platform (Figure 1a).[2a] In this technology, a proteome of interest is split into two samples. One of these samples is treated with a covalent inhibitor and the other one with DMSO as a control. In the next step, both samples are treated with iodoacetamide alkyne (IA-alkyne).[9] This probe will modify many cysteines in both samples with alkynes and this reactivity will be blocked by the covalent inhibitor at its specific binding sites. The samples are next modified with isotopically labeled affinity tags using copper-catalyzed azide–alkyne cycloaddition (CuAAC).[10] The samples are combined, enriched on streptavidin beads, proteolytically digested and the modified peptides eluted for mass spectrometry (MS) based quantification. Most quantified cysteines will have ratios $R \approx 1$ between the heavy and light channel indicating no interaction with the covalent compound (Figure 1a). In contrast, cysteines at the specific binding sites will show ratios of $R \gg 1$. In this way, quantitative and site-specific interaction studies in the whole proteome are possible with unmodified covalent inhibitors that do not need to be equipped, for example, with an affinity handle.

In the last step of this protocol, the modified peptides need to be eluted from the streptavidin beads for MS-based analysis. As previous studies have utilized biotin, which binds almost irreversibly to streptavidin, as an affinity handle, various cleavable linkers have been applied to elute the peptides from the beads.[11] These linkers include those that are cleaved by proteases (Figure 1b),[2a,9] acidic,[11b,12] or reductive conditions.[11b,13] Because of the high requirements on the orthogonality of these linkers, they need to be designed very carefully, which usually requires laborious multistep synthesis of the tags. Furthermore, the cleavage of the linker adds another step to the chemoproteomic protocol.

Therefore, we set out to develop isotopically labeled desthiobiotin azide (isoDTB) tags (Figure 1c) for residue-

[*] M. Sc. P. R. A. Zanon, M. Sc. L. Lewald, Dr. S. M. Hacker
Department of Chemistry, Technical University of Munich
Lichtenbergstrasse 4, 85748 Garching (Germany)
E-mail: stephan.m.hacker@tum.de

[**] A previous version of this manuscript has been deposited on a preprint server (https://doi.org/10.26434/chemrxiv.9853445.v1).

Supporting information and the ORCID identification number(s) for the author(s) of this article can be found under: https://doi.org/10.1002/anie.201912075.

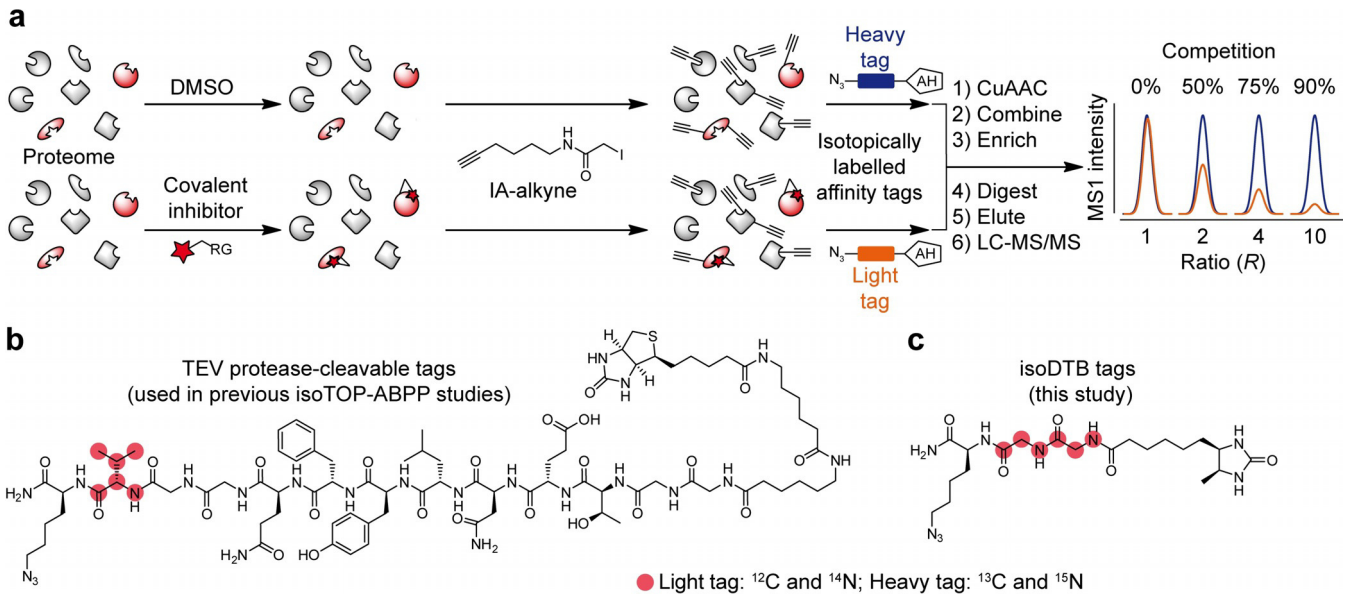


Figure 1. a) Workflow for competitive, residue-specific chemoproteomic experiments.[2a] RG = reactive group, AH = affinity handle. b) Structure of the TEV protease-cleavable tags (TEV tags) originally used for residue-specific proteomics.[2a,9] c) Structure of the isoDTB tags developed in this study.

specific proteomics. As desthiobiotin still binds very strongly to streptavidin, all steps up to the proteolytic digestion can be kept the same.[14] Because of the reversibility of binding of desthiobiotin to streptavidin, in the last step, the peptides are then easily eluted using acidic conditions with acetonitrile as the cosolvent.[14] Because a complex cleavable linker is not needed, we designed these tags exclusively with two isotopically differentiated glycine residues as the linker moiety.

After establishing the utility of the isoDTB tags for residue-specific proteomics, we used them to globally investigate cysteines in the proteome of *S. aureus* for their reactivity and their potential to bind covalent ligands. In this way, we identified 88 highly reactive cysteines and more than 250 cysteines that can be addressed with covalent ligands. These residues are starting points for the development of antibiotics with novel modes-of-action.

Results and Discussion

We synthesized the isoDTB tags using solid-phase peptide synthesis. For this purpose, a Rink amide resin and an Fmoc strategy were utilized. We sequentially coupled ε-azido-lysine, two glycine residues, and desthiobiotin. We used glycine with the natural isotope distribution for the light isoDTB tag and glycine with two ^{13}C atoms and one ^{15}N atom for the heavy tag. In this way, a total mass difference between the tags of 6 Da was obtained. Purification by RP-HPLC resulted in a yield of approximately 70 % for both isoDTB tags.

To establish that the tags are applicable to broadly investigate cysteines in a proteomic context, we treated two identical samples of the lysate of the methicillin-sensitive *S. aureus* (MSSA) strain SH1000[15] with 1 mM IA-alkyne and modified the two samples with the light and heavy isoDTB tag, respectively, using CuAAC. The samples were combined either in a ratio of 1:1 or 1:4. Subsequently, we enriched the samples on streptavidin beads, digested the proteins with trypsin, and eluted the modified peptides using our straightforward approach. Analysis using liquid chromatography coupled to tandem MS (LC-MS/MS) using a Q Exactive Plus (Thermo Fisher) mass spectrometer and evaluation using freely available MaxQuant software[16] identified 1155 cys-

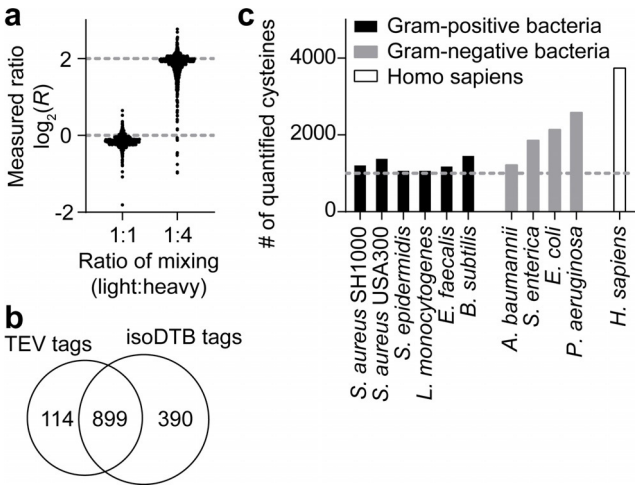


Figure 2. a) Ratios *R* of all quantified cysteines in the *S. aureus* SH1000 proteome in experiments, in which the light and heavy labeled samples were both reacted with 1 mM IA-alkyne, clicked to the isoDTB tags and mixed at the indicated ratios. Expected values of $\log_2(R)$ of 0 for the 1:1 mixture and 2 for the 1:4 mixture are indicated with dashed grey lines. b) Venn diagram comparing the number of quantified cysteines in the *S. aureus* SH1000 proteome using 1 mM IA-alkyne and the TEV tags or the isoDTB tags, respectively. c) Number of quantified cysteines in a variety of Gram-positive and Gram-negative bacteria as well as in the human cell line MDA-MB-231 using the isoDTB tags. The grey dashed line indicates 1000 quantified cysteines. All data results from duplicates.

teines that were quantified for both conditions (Figure 2a; see Table S1). This analysis revealed a narrow distribution of the detected ratios for both samples around the expected values. The isoDTB tags therefore reliably allowed quantification of cysteines in the whole bacterial proteome.

We next benchmarked our technology against the TEV protease-cleavable biotin tags (TEV tags) that have been most broadly used to residue-specifically map proteomes (Figure 2b; see Table S1).[2a,9] Our isoDTB tags outperformed the TEV tags by quantifying 27% more cysteines in the *S. aureus* proteome. We increased the number of cysteines quantified with the isoDTB tags even more by additionally using chymotrypsin and AspN for the proteolytic digest (see Figure S1 and Table S1). These experiments are not possible for the TEV tags as these proteases would cleave the tag itself. In this way, we quantified a total of 1643 cysteines in the *S. aureus* proteome using the isoDTB tags in only six experiments. We next investigated the performance of the isoDTB tags with IA-alkyne in different Gram-positive and Gram-negative bacteria (Figure 2c; see Table S1), and consistently quantified more than 1000 cysteines in each strain. Moreover, we were able to quantify more than 3500 cysteines in the human cell line MDA-MB-231, which is competitive with previously described methods.[2a] Therefore, our isoDTB tags not only shortened the chemoproteomic protocol but also led to increased coverage in bacterial systems compared to the widely used TEV tag technology.

We next applied our method to analyze the reactivity of cysteines in the bacterial proteome (Figure 3a).[9a] As the reactivity of cysteines is linked to their functional relevance in human cells,[9a] we reasoned that this feature might also be conserved in bacteria and in this way lead to the identification of functionally important cysteine residues. To study cysteine reactivity, two identical samples of the proteome of the *S. aureus* strain SH1000 were treated with either a high (100 μM) or a low (10 μM) concentration of IA-alkyne. In this way, while at the high concentration many cysteines were labeled, at the low concentration only the most reactive cysteines were labeled quantitatively. After CuAAC with the light (low concentration) and heavy isoDTB tags (high concentration), respectively, the samples were analyzed in the same way as described above. Here, high ratios ($R_{10:1}$) indicate low reactivity cysteines, whereas the most reactive cysteines have $R_{10:1} \approx 1$. Using this procedure, we quantified 921 cysteines and identified 88 highly reactive cysteines with $R_{10:1} < 3$ in 69 different proteins (Figure 3b; see Table S2). Another 240 cysteines showed medium reactivity ($3 < R_{10:1} < 5$), whereas the remaining 593 cysteines were of low reactivity ($R_{10:1} > 5$). Cysteines of all three bins of reactivity were evenly distributed throughout the different functional classes of proteins (see Figure S2).[18] Interestingly, highly reactive cysteines were depleted in essential proteins[17] in comparison to their counterparts of lower reactivity (Figure 3c). It can be speculated that evolutionary pressure has selected against highly reactive cysteines in essential proteins as these would interact with many reactive small-molecule electrophiles that occur in nature.

There is a strong enrichment of the highly and medium reactive cysteines at functional sites (Figure 3d). These cysteines include many residues that are directly involved in the catalytic mechanism (e.g. C178 in the GTP cyclohydrolase FolE2 (UniProt code Q2G0L1),[19] C112 in FabH (UniProt

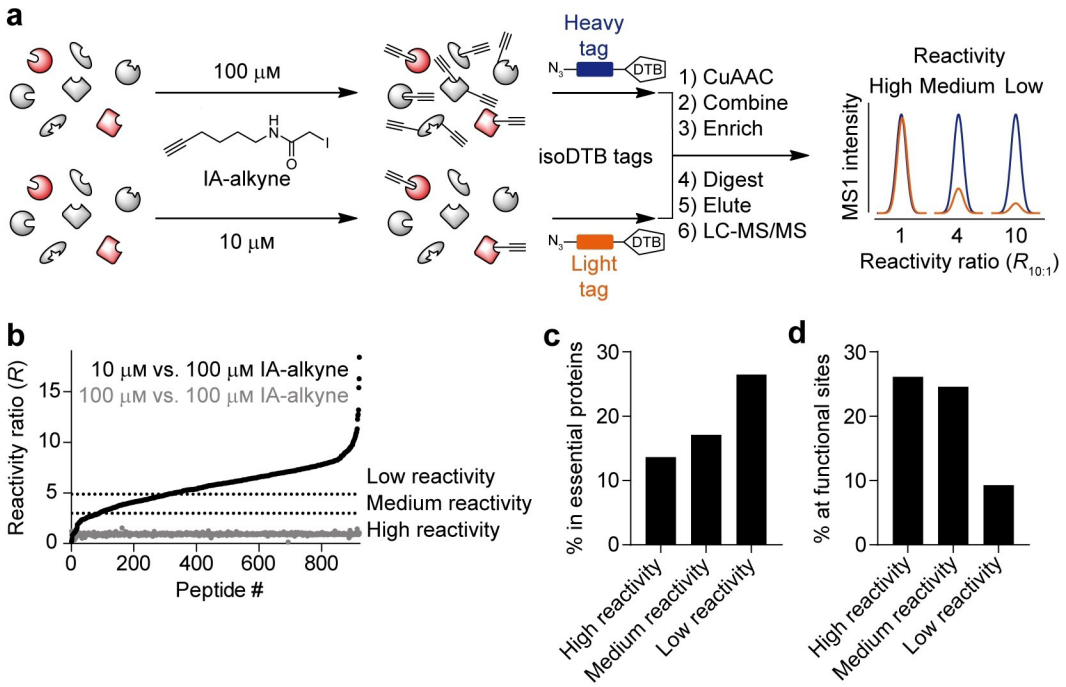


Figure 3. a) Workflow for the measurement of cysteine reactivity using the isoDTB tags.[9a] DTB = desthiobiotin. b) Plot of the reactivity ratios ($R_{10:1}$) obtained by comparing *S. aureus* SH1000 proteomes treated with high (100 μM) *vs.* low concentration (10 μM) of IA-alkyne (black). Ratios ($R_{1:1}$) of an experiment with high concentration used for both samples (grey) are used as a control to ensure reliable quantification of all cysteines. c,d) Percentage of cysteines in the different reactivity bins that are in essential proteins (c)[17] or at functional sites (d).

code Q2FZS0), and C88 in the probable acetyl-CoA acyltransferase (UniProt code Q2G124), which are all essential proteins). Furthermore, several highly reactive cysteines are close to cofactor-binding sites (e.g. C239 of the CTP synthase PyrG (UniProt code Q2FWD1), and C45 in MnmG (UniProt code Q2FUQ3)) or metal-binding sites (e.g. C145 of alcohol dehydrogenase Adh (UniProt code Q2G0G1), and C65 of biotin synthase BioB (UniProt code Q2FVJ7)). Therefore, residue-specific proteomics using our isoDTB tags allowed global profiling of the reactivity of cysteines in the bacterial proteome and enabled the identification of functionally relevant residues.

We next set out to study which cysteines in the *S. aureus* proteome can be targeted with covalent ligands. For this purpose, we obtained a library of 211 electrophilic cysteine-directed compounds (**EN001**–**EN211**; see Table S3), mainly α-chloroacetamides. These compounds were initially screened for antibacterial activity by performing minimum inhibitory concentration (MIC) experiments. While we did not expect these small compounds to be completely specific, we used this phenotypic pre-filter to prioritize compounds, whose target spectrum includes essential proteins that can be addressed in intact cells. Based on an initial screen in three MSSA strains, we selected 24 compounds (see Figure S3) based on their MIC values and structural diversity for further studies. Interestingly, many of these compounds contain a 2-aminothiazole moiety, which seems to be beneficial for activity. 23 compounds had MICs of ≤ 100 μM in all three strains with six compounds having MICs of ≤ 12.5 μM in all three strains (see Figure S4). Furthermore, 14 compounds showed activity (MIC ≤ 100 μM) in two tested MRSA strains with two compounds (**EN085** and **EN177**) having an MIC ≤ 10 μM in all five tested strains. This data shows that electrophilic compounds with desired biological activity could efficiently be identified from a small compound library.

The selected 24 compounds were screened at a 200 μM concentration in residue-specific chemoproteomic experiments using our isoDTB tags in duplicates (Figure 1a; see Figure S5 and Figure S6). For three of the compounds (**EN007**, **EN085** and **EN177**), we performed an additional set of biologically independent duplicates. Given the high reproducibility between the biologically independent experiments (see Figure S5), we performed the remaining profiling in duplicates and prioritized screening more compounds over performing more replicates. Five of the compounds (**EN007**, **EN085**, **EN135**, **EN177**, and **EN201**) that showed MIC values ≤ 25 μM in all five tested strains were additionally tested at 20 μM concentration (see Figure S7).

In all experiments, we consider cysteines that have a ratio of $R > 4$ ($\log_2(R) > 2$) and whose R value is statistically significantly different from $R = 1$ (p-value < 0.05 in a one-sample t-test), to be engaged by the covalent ligand. We identified a large range of values for the fraction of cysteines that are engaged by the different compounds (Figures 4a–c; see Figure S6 and Figure S8a). Nine compounds showed low promiscuity (< 2% of all quantified cysteines are engaged, Figure 4a), ten compounds showed medium promiscuity (2% to 10%, Figure 4b), and five compounds showed excessive promiscuity (> 10%, Figure 4c). Strikingly, no correlation between MIC and promiscuity could be observed (see Figure S8b), indicating that it is possible to identify highly active and still selective electrophiles. As we cannot rule out unspecific effects for the highly promiscuous compounds, we excluded these from all further analysis. While the low promiscuity compounds are most interesting for further compound development, the medium promiscuity compounds are most useful for the global profiling approach performed here.

Taking all 25 investigated conditions together (19 compounds at 200 μM, five compounds at 20 μM, and a DMSO control), we compiled a competitive data table (see Table S4), which includes all cysteines that were quantified for at least three of the conditions. In this way, we obtained information on 1756 cysteines in 905 different proteins, which corresponds to a coverage of 33% of all the cysteines encoded in the *S. aureus* genome. As cysteines in essential proteins[17] are enriched in our data over the genomic background (see Figure S9a), this equates to the quantification of 59% of all cysteines in essential proteins. Each cysteine was quantified on average for 21 of the 25 conditions (see Figure S10). Therefore, our method allowed obtaining information on many cysteine residues in *S. aureus* in a reproducible manner.

268 cysteines in 200 different proteins were engaged by at least one ligand (Figure 5a). In many proteins, we detect one

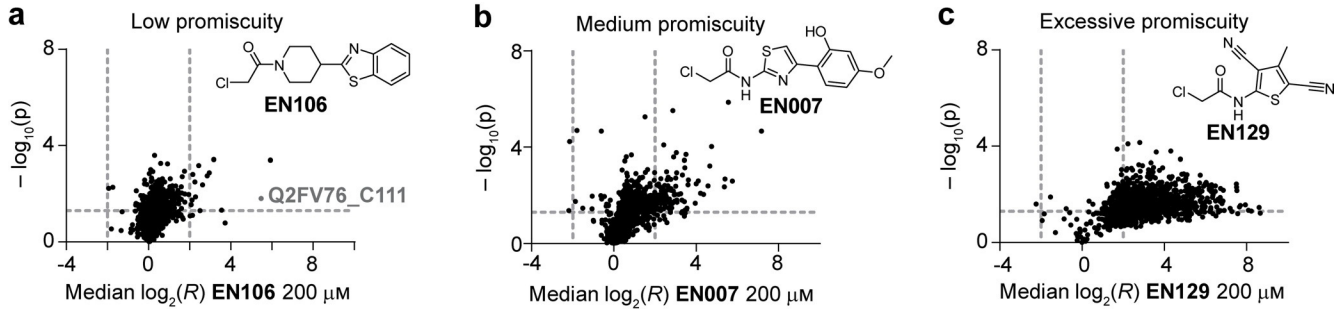


Figure 4. a–c) Volcano plots for three representative compounds of low (a), medium (b) and excessive promiscuity (c). Plots show the $\log_2(R)$ of the ratio between the heavy (DMSO-treated) and light (compound-treated) channels and the $-\log_{10}(p)$ of the statistical significance in a one-sample t-test for all quantified cysteines. In plot a) the data point for the ligandable active site residue C111 of the essential putative HMG-CoA synthase (UniProt code Q2FV76) is highlighted in grey. All data results from duplicates. For **EN007** an additional set of biologically independent duplicates was included.

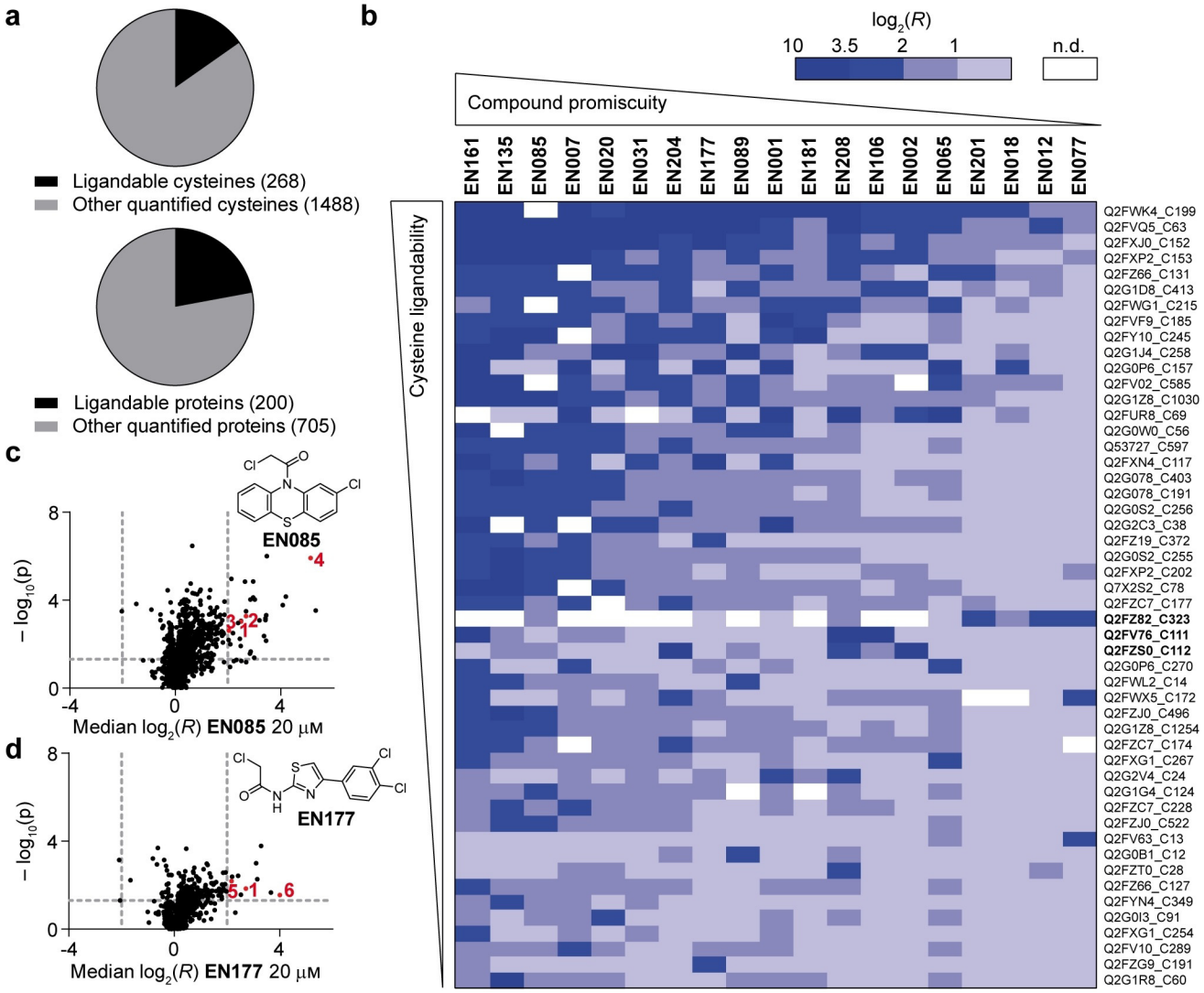


Figure 5. a) Total number of ligandable and other quantified cysteines and proteins in our competitive data table. b) Heat map of the $\log_2(R)$ values for a selection of ligandable cysteines with all tested compounds at 200 µM. Cysteines discussed in the text are highlighted in boldface. Compounds are sorted from left to right by decreasing promiscuity and cysteines are sorted from top to bottom by decreasing number of identified ligands. c,d) Volcano plots for compounds **EN085** (c) and **EN177** (d) at 20 µM in *S. aureus* SH1000. Plots show the $\log_2(R)$ of the ratio between the heavy (DMSO-treated) and light (compound-treated) channels and the $-\log_{10}(p)$ of the statistical significance in a one-sample t-test for all quantified cysteines. Ligandable cysteines in essential proteins that are engaged by the respective compound are highlighted in red. 1: C152 of MurC (UniProt code Q2FXJ0), 2: C199 of MnmA (UniProt code Q2FXV6), 3: C1030 of PolC (UniProt code Q2G1Z8), 4: C151 of glyceraldehyde-3-phosphate dehydrogenase (UniProt code Q2G032), 5: C410 of pyruvate kinase (UniProt code Q2FXM9), 6: C88 of probable acetyl-CoA acyltransferase (UniProt code Q2G124). All data results from duplicates. For compounds **EN007**, **EN085** and **EN177** in panel b) and all data in panels c) and d) an additional biologically independent set of duplicates was also included.

ligandable cysteine that is engaged by several compounds, while the other cysteine(s) are never engaged (see Figure S11). This observation indicates that our method measures local target engagement of the cysteines rather than global changes to the protein structure. While ligandable cysteines are enriched in enzymes, we also identify them in other functional classes of proteins (see Figure S12).[18] Ligandable cysteines are enriched at functional sites (see Figure S13) and similarly abundant in essential proteins as compared to other quantified cysteines (see Figure S9b). When we compared this data to the cysteine reactivity data (see Figure S14), we could see that, while the highly reactive cysteines are clearly more likely to be ligandable, there are also many ligandable cysteines of medium and low reactivity, indicating that specific noncovalent interactions are important in these cases.

Next, we looked at the binding of our covalent ligands to a selection of ligandable cysteines in more detail (Figure 5b). While the most ligandable cysteines tend to be engaged by the most promiscuous compounds, there is clear evidence for more specific interactions between less ligandable cysteines with more selective compounds. For example, the active site residue C112 of FabH (UniProt code Q2FZS0), an essential enzyme in fatty acid synthesis,[20] is exclusively targeted by

three compounds of tempered promiscuity (**EN002**, **EN204** and **EN208**, Figure 5b). This residue has previously been shown to be covalently modified for example, by the inhibitors 4,5-dichloro-1,2-dithiol-3-one and cerulenin.[21] Furthermore, the residue C323 in the isoleucine-tRNA ligase IleS (UniProt code Q2FZ82) is only targeted by another set of three compounds (**EN012**, **EN077** and **EN201**, Figure 5b), which could open up the possibility of inhibiting bacterial translation through a novel target. Overall, we detect many binding events that are strongly dependent on the compound and on the targeted cysteine indicating that our method can detect specific ligand-binding events.

Looking at the targets of the two compounds that showed the best antibacterial activity in the initial MIC assays (**EN085** and **EN177**, Figure 5c,d), we saw that both compounds show engagement of several cysteines at 20 μM (31 for **EN085**, 10 for **EN177**). Both compounds strongly target C152 of MurC (UniProt code Q2FXJ0), which is a key enzyme essential for cell wall synthesis.[22] **EN177** additionally binds to C410 of pyruvate kinase (UniProt code Q2FXM9) and C88 of the essential probable acetyl-CoA acyltransferase (UniProt code Q2G124). The latter cysteine forms an acyl-thioester intermediate during catalysis.[23] **EN085** binds to C1030 in the DNA polymerase PolC (UniProt code Q2G1Z8), C199 in the tRNA-specific methyl transferase MnmA (UniProt code Q2FXV6), which forms a cysteine persulfide intermediate during catalysis,[23] as well as the catalytically active nucleophile C151 in glyceraldehyde-3-phosphate dehydrogenase (UniProt code Q2G032).[24] Both compounds, therefore, bind to several essential target proteins that have the potential to become novel targets of covalent antibiotic compounds.

To investigate if the results obtained in the MSSA strain SH1000 are transferable to other *S. aureus* strains, compounds **EN085** and **EN177** were additionally screened at 20 μM in the MRSA strain USA300, for which they show MIC values of 6.3 μM and 3.1 μM, respectively (see Figure S15 and Table S5). We detect a very good correlation of the data obtained in the two different strains (see Figure S16). All cysteines in essential proteins discussed above were also engaged by the same compound in USA300. While no new engaged cysteines were identified for **EN177**, we identified five additional engaged cysteines for **EN085** in USA300 that were not quantified at all in SH1000. Among those, two cysteines are in essential proteins.[17] **EN085** binds to the active site C119 in MurA (UniProt code A0A0H2XGP3), which is a key enzyme in cell wall biosynthesis[22] and also targeted by fosfomycin.[5] Additionally, C565 in the aspartate-tRNA ligase AspS (UniProt code Q2FG97) is modified by **EN085**, which opens up the possibility to target translation through a novel mechanism. The highly reproducible results between the MSSA and MRSA strains demonstrate that our data delivers a broadly applicable map of ligandable cysteines in the *S. aureus* proteome that will guide the design of antibiotics with novel modes-of-action.

To validate the interaction of a selected compound with an identified ligandable cysteine, we investigated C111 of the putative HMG-CoA synthase (UniProt code Q2FV76), which is an essential enzyme in the mevalonate pathway and in this way might open up targeting bacteria through this so far clinically unexplored pathway.[25] In gel-based experiments (Figure 6a; see Figure S17), strong labeling by IA-alkyne was observed for the recombinant wildtype protein, but not for the C111A mutant. This data is in good agreement with the high reactivity of C111 ($R_{10:1} = 0.84$) in the reactivity experiments (Figure 3; see Table S2). Furthermore, the low promiscuity compound **EN106** that we identified to target HMG-CoA synthase (Figure 4a) blocked labeling at low micromolar concentrations, indicating covalent binding of this compound to C111. Using intact protein MS (IPMS, Figure 6b; see Figures S18 and S19), we detected quantitative single modification of the HMG-CoA synthase wildtype with **EN106**. No modification of the C111A mutant was detectable,

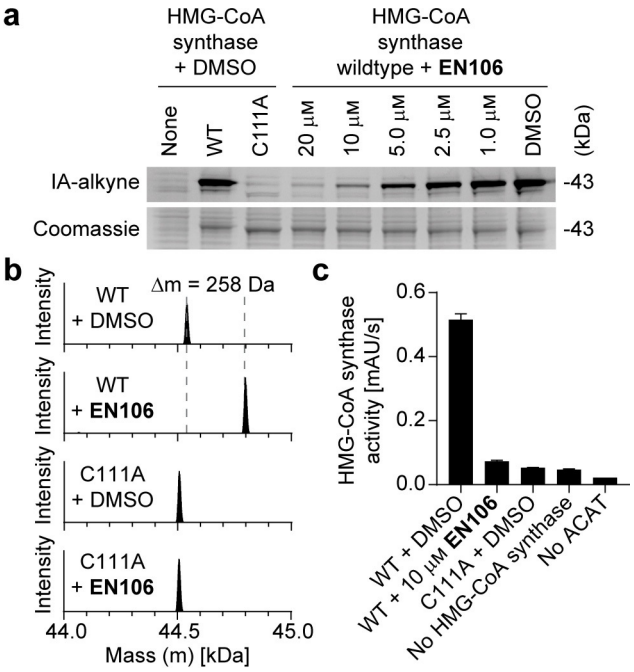


Figure 6. a) Result of gel-based labeling experiments with HMG-CoA synthase. 1 μM recombinant wild-type (WT) HMG-CoA synthase was added into 1 mg mL^{-1} soluble lysate of *S. aureus* SH1000. As controls, 1 μM of the HMG-CoA synthase mutant (C111A) or no HMG-CoA synthase (none) were added. The samples were treated with the indicated concentrations of **EN106** or with DMSO as control. The samples were labeled with IA-alkyne and modified with TAMRA-azide using CuAAC. Analysis using SDS-PAGE with subsequent in-gel fluorescence scanning and Coomassie staining is shown. b) IPMS analysis of the modification of HMG-CoA synthase by **EN106**. 1 μM HMG-CoA synthase wildtype (WT) or mutant (C111A) was treated with DMSO as control or 10 μM **EN106**. Deconvoluted IPMS spectra are shown. The mass difference between the wildtype treated with **EN106** or DMSO ($\Delta m = 258$ Da) exactly corresponds to the modification of the protein with one molecule of **EN106**. c) Results of activity assays with HMG-CoA synthase. 1 μM HMG-CoA synthase wildtype (WT) was treated with 10 μM **EN106** or DMSO as a control. Acetyl-CoA, acetyl-CoA-acyl-transferase (ACAT) and Ellman's reagent were added and the reaction progress was followed by measuring the absorbance at 410 nm over time. HMG-CoA synthase activity was calculated by a linear fit of the linear portion of this curve. Controls with the HMG-CoA synthase mutant (C111A), no HMG-CoA synthase or no acetyl-CoA-acyl-transferase (no ACAT) were included. The graph shows mean ± standard deviation. All data results from triplicates. mAU: milli absorbance units.

strongly indicating that C111 is the site of covalent modification with **EN106**. To study the activity of HMG-CoA synthase (Figure 6c; see Figure S20) we set up a coupled assay with acetyl-CoA-acyl transferase (ACAT).[26] ACAT forms acetoacetyl-CoA from two molecules of acetyl-CoA. HMG-CoA synthase then catalyzes the reaction with another molecule of acetyl-CoA to give HMG-CoA. The free thiol group of the CoA-SH liberated in both steps was detected using Ellman's reagent (Figure 6c). Addition of wild-type HMG-CoA synthase to the assay strongly increased the formation of free CoA-SH over the ACAT background reaction. This activity was reduced to the level without HMG-CoA synthase, when the C111A mutant was used or when the wildtype was pretreated with 10 μM **EN106. EN106** had no effect on the detected ACAT activity in absence of HMG-CoA synthase or presence of the inactive C111A mutant (see Figure S21), showing that **EN106** does not inhibit the ACAT reaction or hinder detection by alkylation of the free CoA-SH or the product of the Ellman's reagent. Furthermore, inhibition remained after gel-filtration to remove excess free **EN106**, showing irreversible inhibition and further excluding interference of **EN106** with other components of the assay (see Figure S22). Covalent modification of HMG-CoA synthase at C111 with compound **EN106** therefore led to effective inhibition of its activity. HMG-CoA synthase is therefore a promising target for the development of novel antibiotics that interfere with the essential mevalonate pathway.

Conclusion

We describe the synthesis of isotopically labeled desthiobiotin azide (isoDTB) tags and their application in chemoproteomic experiments. These tags were easily synthesized by solid-phase peptide synthesis in high yields and showed excellent physicochemical properties. By using desthiobiotin, these tags circumvented the need to use complex cleavable linkers[11a] for peptide elution and thus significantly shortened the chemoproteomic protocol, while increasing the coverage of cysteines in the proteome of *S. aureus*. The isoDTB tags allowed quantification of many cysteines across different Gram-positive and Gram-negative bacterial proteomes and gave results comparable to the TEV tags, also in the human proteome.[2a] Because of the easy synthesis of the tag, the shortened workflow, the use of freely available MaxQuant data evaluation software,[16] and the excellent performance, this technology will make residue-specific proteomics applicable in many laboratories not specialized in chemoproteomics.

The isoDTB tags were applied to study the reactivity of cysteines in the proteome of *S. aureus*. We identified 88 highly reactive cysteine residues that are strongly enriched at functional sites of proteins. This enrichment indicates that the reactivity of cysteines is a proxy for the functional relevance of certain residues also in bacterial proteomes. Interestingly, highly reactive cysteines were less likely to be found in essential proteins, pointing to the fact that evolution may have selected against highly reactive cysteines in essential proteins to protect bacteria from the influence of reactive electrophiles occurring in nature either during metabolism or as environmental chemicals.

Finally, we applied the isoDTB tags to broadly understand which cysteines in the bacterial proteome can be engaged with covalent ligands. For this purpose, we compiled competitive data for 19 α-chloroacetamides and profiled 1756 cysteines, including 59% of all cysteines in essential proteins. We identified 268 cysteines that can bind covalent ligands in 200 different proteins. The targeted cysteines include many functionally relevant residues in essential proteins involved in many different pathways. In this way, the data presented will be the starting point for more specific covalent inhibitors to develop antibiotics with novel modes-of-action. The presented isoDTB tags will allow monitoring of the on- and off-target effects of the compounds and in this way streamline the development process.

We investigated inhibition of HMG-CoA synthase in more detail. Modification at the ligandable cysteine residues was detected and this interaction led to inhibition of the enzyme activity. HMG-CoA synthase inhibition by modification of C111 using the human HMG-CoA synthase inhibitor hymeglusin has been described, but this inhibitor suffers from a short half-life of the thioester in the covalent protein adduct.[27] Therefore, permanent covalent inhibition by the low promiscuity compound **EN106** is a very promising starting point to explore the antibiotic potential of this protein.[25] This case study demonstrates that our map of ligandable cysteines is an excellent resource to quickly identify residues that can be targeted in a functionally relevant manner.

Taken together, our isoDTB tags are important new tools for residue-specific proteomics in bacterial systems. They allowed the investigation of the bacterial cysteinome globally and should be transferable to studying other amino acids in a straightforward manner.[8] The cysteines that were characterized to bind to covalent ligands in this study serve as the foundation for the development of covalent inhibitors that could lead to antibiotics with totally new modes-of-action.

Acknowledgements

We gratefully acknowledge Prof. Dr. Stephan A. Sieber and his group for their generous support. We thank Dr. Markus Lakemeyer and Dr. Franziska Mandl for critical reading of the manuscript. SMH and PRAZ acknowledge funding by the Fonds der Chemischen Industrie through a Liebig Fellowship and a Ph.D. fellowship. We acknowledge financial support by the TUM Junior Fellow Fund.

Conflict of interest

The authors declare no conflict of interest.

Keywords: antibiotics · covalent inhibitors · isotopic labeling · protein modifications · proteomics

How to cite: *Angew. Chem. Int. Ed.* **2020**, *59*, 2829 – 2836
Angew. Chem. **2020**, *132*, 2851 – 2858

[1] M. Lakemeyer, W. Zhao, F. A. Mandl, P. Hammann, S. A. Sieber, *Angew. Chem. Int. Ed.* **2018**, *57*, 14440 – 14475; *Angew. Chem.* **2018**, *130*, 14642 – 14682.
[2] a) K. M. Backus, B. E. Correia, K. M. Lum, S. Forli, B. D. Horning, G. E. Gonzalez-Paez, S. Chatterjee, B. R. Lanning, J. R. Teijaro, A. J. Olson, D. W. Wolan, B. F. Cravatt, *Nature* **2016**, *534*, 570 – 574; b) C. G. Parker, A. Galmozzi, Y. Wang, B. E. Correia, K. Sasaki, C. M. Joslyn, A. S. Kim, C. L. Cavallaro, R. M. Lawrence, S. R. Johnson, I. Narvaiza, E. Saez, B. F. Cravatt, *Cell* **2017**, *168*, 527 – 541.e529; c) C. Y. Chung, H. R. Shin, C. A. Berdan, B. Ford, C. C. Ward, J. A. Olzmann, R. Zoncu, D. K. Nomura, *Nat. Chem. Biol.* **2019**, *15*, 776 – 785; d) C. A. Berdan, R. Ho, H. S. Lehtola, M. To, X. Hu, T. R. Huffman, Y. Petri, C. R. Altobelli, S. G. Demeulenaere, J. A. Olzmann, T. J. Maimone, D. K. Nomura, *Cell Chem. Biol.* **2019**, *26*, 1027 – 1035.e1022.
[3] J. Singh, R. C. Petter, T. A. Baillie, A. Whitty, *Nat. Rev. Drug Discovery* **2011**, *10*, 307 – 317.
[4] A. Chaikuad, P. Koch, S. A. Laufer, S. Knapp, *Angew. Chem. Int. Ed.* **2018**, *57*, 4372 – 4385; *Angew. Chem.* **2018**, *130*, 4456 – 4470.
[5] D. Hendlin, E. O. Stapley, M. Jackson, H. Wallick, A. K. Miller, F. J. Wolf, T. W. Miller, L. Chaiet, F. M. Kahan, E. L. Foltz, H. B. Woodruff, J. M. Mata, S. Hernandez, S. Mochales, *Science* **1969**, *166*, 122 – 123.
[6] H. Nishimura, M. Mayama, Y. Komatsu, H. Kato, N. Shimaoka, Y. Tanaka, *J. Antibiot.* **1964**, *17*, 148 – 155.
[7] P. A. Smith et al., *Nature* **2018**, *561*, 189 – 194.
[8] S. M. Hacker, K. M. Backus, M. R. Lazear, S. Forli, B. E. Correia, B. F. Cravatt, *Nat. Chem.* **2017**, *9*, 1181 – 1190.
[9] a) E. Weerapana, C. Wang, G. M. Simon, F. Richter, S. Khare, M. B. Dillon, D. A. Bachovchin, K. Mowen, D. Baker, B. F. Cravatt, *Nature* **2010**, *468*, 790 – 795; b) A. E. Speers, B. F. Cravatt, *J. Am. Chem. Soc.* **2005**, *127*, 10018 – 10019.
[10] V. V. Rostovtsev, L. G. Green, V. V. Fokin, K. B. Sharpless, *Angew. Chem. Int. Ed.* **2002**, *41*, 2596 – 2599; *Angew. Chem.* **2002**, *114*, 2708 – 2711.
[11] a) Y. Yang, M. Fonovic, S. H. Verhelst, *Methods Mol. Biol.* **2017**, *1491*, 185 – 203; b) A. J. Rabalski, A. R. Bogdan, A. Baranczak, *ACS Chem. Biol.* **2019**, *14*, 1940 – 1950.
[12] J. Szychowski, A. Mahdavi, J. J. Hodas, J. D. Bagert, J. T. Ngo, P. Landgraf, D. C. Dieterich, E. M. Schuman, D. A. Tirrell, *J. Am. Chem. Soc.* **2010**, *132*, 18351 – 18360.
[13] a) S. H. Verhelst, M. Fonovic, M. Bogyo, *Angew. Chem. Int. Ed.* **2007**, *46*, 1284 – 1286; *Angew. Chem.* **2007**, *119*, 1306 – 1308; b) Y. Qian, J. Martell, N. J. Pace, T. E. Ballard, D. S. Johnson, E. Weerapana, *ChemBioChem* **2013**, *14*, 1410 – 1414.
[14] M. P. Patricelli et al., *Chem. Biol.* **2011**, *18*, 699 – 710.
[15] M. J. Horsburgh, J. L. Aish, I. J. White, L. Shaw, J. K. Lithgow, S. J. Foster, *J. Bacteriol.* **2002**, *184*, 5457 – 5467.
[16] J. Cox, M. Mann, *Nat. Biotechnol.* **2008**, *26*, 1367 – 1372.
[17] R. R. Chaudhuri, A. G. Allen, P. J. Owen, G. Shalom, K. Stone, M. Harrison, T. A. Burgis, M. Lockyer, J. Garcia-Lara, S. J. Foster, S. J. Pleasance, S. E. Peters, D. J. Maskell, I. G. Charles, *BMC Genomics* **2009**, *10*, 291.
[18] P. Gaudet, M. S. Livstone, S. E. Lewis, P. D. Thomas, *Briefings Bioinf.* **2011**, *12*, 449 – 462.
[19] N. Paranagama, S. A. Bonnett, J. Alvarez, A. Luthra, B. Stec, A. Gustafson, D. Iwata-Reuyl, M. A. Swairjo, *Biochem. J.* **2017**, *474*, 1017 – 1039.
[20] J. Yao, C. O. Rock, *Biochim. Biophys. Acta Mol. Cell Biol. Lipids* **2017**, *1862*, 1300 – 1309.
[21] a) J. B. Parsons, J. Yao, M. W. Frank, C. O. Rock, *Antimicrob. Agents Chemother.* **2015**, *59*, 849 – 858; b) A. G. Ekström, V. Kelly, J. Marles-Wright, S. L. Cockroft, D. J. Campopiano, *Org. Biomol. Chem.* **2017**, *15*, 6310 – 6313.
[22] T. D. Bugg, D. Braddick, C. G. Dowson, D. I. Roper, *Trends Biotechnol.* **2011**, *29*, 167 – 173.
[23] The UniProt Consortium, *Nucleic Acids Res.* **2019**, *47*, D506 – D515.
[24] S. Mukherjee, D. Dutta, B. Saha, A. K. Das, *J. Mol. Biol.* **2010**, *401*, 949 – 968.
[25] N. Campobasso, M. Patel, I. E. Wilding, H. Kallender, M. Rosenberg, M. N. Gwynn, *J. Biol. Chem.* **2004**, *279*, 44883 – 44888.
[26] D. A. Skaff, H. M. Miziorko, *Anal. Biochem.* **2010**, *396*, 96 – 102.
[27] D. A. Skaff, K. X. Ramyar, W. J. McWhorter, M. L. Barta, B. V. Geisbrecht, H. M. Miziorko, *Biochemistry* **2012**, *51*, 4713 – 4722.

Manuscript received: September 23, 2019
Accepted manuscript online: November 29, 2019
Version of record online: January 7, 2020

2. Light-Activatable, 2,5-Disubstituted Tetrazoles for the Proteome-Wide Profiling of Aspartates and Glutamates in Living Bacteria

Published in *ACS Central Science* **2020**, *6*, 546-554.

By Kathrin Bach*, Bert L. H. Beerkens*, Patrick R. A. Zanon*, Stephan M. Hacker

https://pubs.acs.org/doi/10.1021/acscentsci.9b01268

(* these authors contributed equally)

Synopsis

As most efforts to develop targeted covalent inhibitors have been directed towards cysteines, we sought to apply our isoDTB-ABPP technology to monitor other proteinogenic amino acids that can be addressed through selective chemistries. Aspartates and glutamates make up around 12% of the bacterial proteome and are commonly found in pockets and protein surfaces due to their high polarity. Furthermore, the unique reactivity of the carboxylic acid side chains could be used to selectively address them.

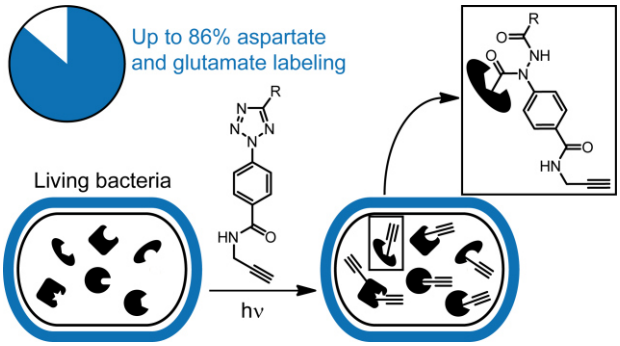
2,5-Disubstituted tetrazoles have been reported to form nitrilimines after UV irradiation. These reactive intermediates can undergo nucleophilic attack by carboxylic acids, followed by an *O*,*N*-acyl shift to give a stable *N*,*N'*-diacylated hydrazine as the product. We synthesized three probes with different residues on the carbon atom of the tetrazole and tested their reactivity and selectivity in the lysate of *S. aureus* using the isoDTB-ABPP platform. The phenyl-substituted probe showed the strongest labeling in comparison with methyl or primary amide analogs but had the lowest selectivity (~60%) for aspartates and glutamates. In contrast, the methyl-probe was more than 85% selective for these residues. Taking the three probes together, we were able to quantify a total of 7995 aspartates and glutamates. We also investigated the reactivity of the methyl-probe on a small molecule level with various amino acids and showed that photolysis of the adduct of the reactive nitrilimine and cysteine could contribute to the observed high selectivity.

Next, we sought to demonstrate that this probe can be used to elucidate the targets of carboxylic acid-directed covalent ligands through competitive residue-specific chemoproteomics. We tested two isoxazolium salts, which have been reported to react with aspartates and glutamates, and identified 44 of these residues as target sites. We also rationalized the use of hydrazonoyl

chlorides as suitable ligands, as they also form nitrilimines through elimination of hydrochloride and thus serve as constitutively reactive counterparts to the light-activated tetrazoles. We synthesized a simple hydrazonoyl chloride, which, among a dozen other aspartates and glutamates, targets D542 of the essential protein nicotinate phosphoribosyltransferase. Although we confirmed this interaction by reductive dimethyl labeling with the recombinantly expressed protein, we found that the enzymatic activity is not affected by the ligand. Nonetheless, this shows that hydrazonoyl chlorides hold promise as a new class of targeted covalent inhibitors. While the aspartates and glutamates addressed by at least one of the three ligands have a slight tendency to be located in functional sites, a larger data set is required for conclusive analysis. For the overall quantified sites, however, essential proteins are strongly enriched over the genomic background.

Since the photoprobes without irradiation are unreactive and non-toxic, they can also be used in living systems. We observed only weak *in situ* labeling with our favored methyl-probe, in stark contrast to the phenyl-probe, which might stem from differences in cell permeation. The phenyl-probe retained its selectivity in this setting and could also be used in two Gram-negative bacterial strains.

In conclusion, the here described tetrazole probes are valuable tools for the monitoring of aspartates and glutamates through residue-specific proteomics both in lysate and in live cells. While we could show target engagement by covalent ligands, not many carboxylic acid-directed chemistries have been reported and we expect our work to foster the development of new targeted covalent inhibitors addressing these amino acids.



Author Contributions

Kathrin Bach., Bert L. H. Beerkens, and Patrick R. A. Zanon contributed equally to this work. Kathrin Bach and Bert L. H. Beerkens performed MS- and gel-based experiments. Kathrin Bach cloned and expressed nicotinate phosphoribosyltransferase and performed biochemical experiments. Bert L. H. Beerkens and Patrick R. A. Zanon synthesized compounds.

Patrick R. A. Zanon performed reactivity assays with amino acids using LC-MS. Kathrin Bach, Bert L. H. Beerkens, and Stephan M. Hacker acquired and analyzed MS data. Stephan M. Hacker conceived and supervised the project. Stephan M. Hacker wrote the manuscript with contributions from all authors.

ACS central science

http://pubs.acs.org/journal/acscii

Research Article

Light-Activatable, 2,5-Disubstituted Tetrazoles for the Proteome-wide Profiling of Aspartates and Glutamates in Living Bacteria

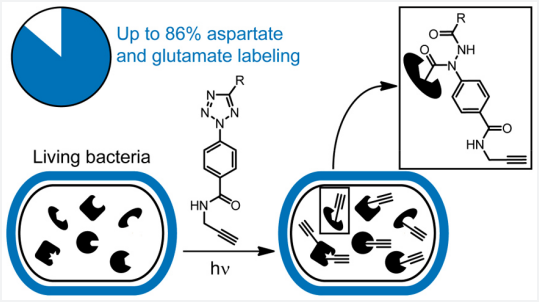
Kathrin Bach,[§] Bert L. H. Beerkens,[§] Patrick R. A. Zanon,[§] and Stephan M. Hacker*

Cite This: *ACS Cent. Sci.* 2020, 6, 546−554

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Covalent inhibitors have recently seen a resurgence of interest in drug development. Nevertheless, compounds, which do not rely on an enzymatic activity, have almost exclusively been developed to target cysteines. Expanding the scope to other amino acids would be largely facilitated by the ability to globally monitor their engagement by covalent inhibitors. Here, we present the use of light-activatable 2,5-disubstituted tetrazoles that allow quantifying 8971 aspartates and glutamates in the bacterial proteome with excellent selectivity. Using these probes, we competitively map the binding sites of two isoxazolium salts and introduce hydrazonyl chlorides as a new class of carboxylic-acid-directed covalent protein ligands. As the probes are unreactive prior to activation, they allow global profiling even in living Gram-positive and Gram-negative bacteria. Taken together, this method to monitor aspartates and glutamates proteome-wide will lay the foundation to efficiently develop covalent inhibitors targeting these amino acids.



INTRODUCTION

Covalent inhibitors have recently re-emerged as important entities in drug development.[1] This is best exemplified by the approval of several kinase inhibitors for clinical use in cancer.[2] Moreover, covalent inhibitors are prevalent among antibiotics. Key examples are the large class of β-lactams[1] but also other antibiotics like fosfomycin,[3] showdomycin,[4] and optimized arylomycins.[5] Nevertheless, covalent inhibitors, which do not rely on an enzymatic activity, still almost exclusively bind to cysteine residues. Targeting additional amino acids could largely help to address protein pockets that do not contain a suitable cysteine and, in this way, enlarge the scope of proteins accessible for covalent inhibitor development.

In the antibiotics field, identifying new binding sites for covalent inhibitors is urgently needed in order to efficiently treat multiresistant bacterial infections.[6] Covalent inhibitors are uniquely suited to identify new targets that can be addressed with small molecules, as they allow efficient mapping of many potential binding sites in parallel using chemoproteomics.[7,8] In bacteria, the almost exclusive focus on cysteine-directed covalent inhibitors raises a severe issue as cysteine is even less frequent in many bacteria (e.g., 0.6% of all amino acid residues in *Staphylococcus aureus* are cysteine) than in human cells (2.3%).[9] Therefore, many important binding pockets in bacterial proteins lack a suitable cysteine residue. Covalent inhibitors that target other amino acid residues would thus be important for antibiotic development, and methods to broadly profile their target engagement with chemoproteomics are highly desirable.

One technology that was key to facilitating the development of covalent inhibitors at cysteines is residue-specific profiling that is usually based on the isoTOP-ABPP (isotopic tandem orthogonal proteolysis activity-based protein profiling) platform (Figure 1A).[7] In this technology, a proteome is split into two samples. One is treated with a covalent inhibitor and the other one with only the solvent as a control. In this way, the inhibitor will covalently bind to its target residues and block their intrinsic reactivity. In the second step, a broadly reactive alkyne probe is used to label many amino acid residues with alkynes. Thereby, binding of the covalent inhibitor is translated into a lack of alkynylation by the probe at the specific interaction sites of the covalent inhibitor. The relative degree of alkynylation in the compound- and solvent-treated samples is quantified by modification with isotopically differentiated affinity tags using copper-catalyzed azide−alkyne cycloaddition (CuAAC). Biotin tags that have an isotopically labeled linker that is cleaved by the tobacco-etch (TEV) protease are most commonly used.[10] Recently, isotopically labeled desthiobiotin azide (isoDTB) tags (Figure 1B) have been introduced that obliviate the need to use a cleavable linker (isoDTB-ABPP).[11] After the combination of the samples, enrichment, and proteolytic digestion, the probe-modified peptides are identified and quantified by liquid chromatography coupled to tandem mass spectrometry

Received: December 11, 2019
Published: April 13, 2020

https://dx.doi.org/10.1021/acscentsci.9b01268
ACS Cent. Sci. 2020, 6, 546−554

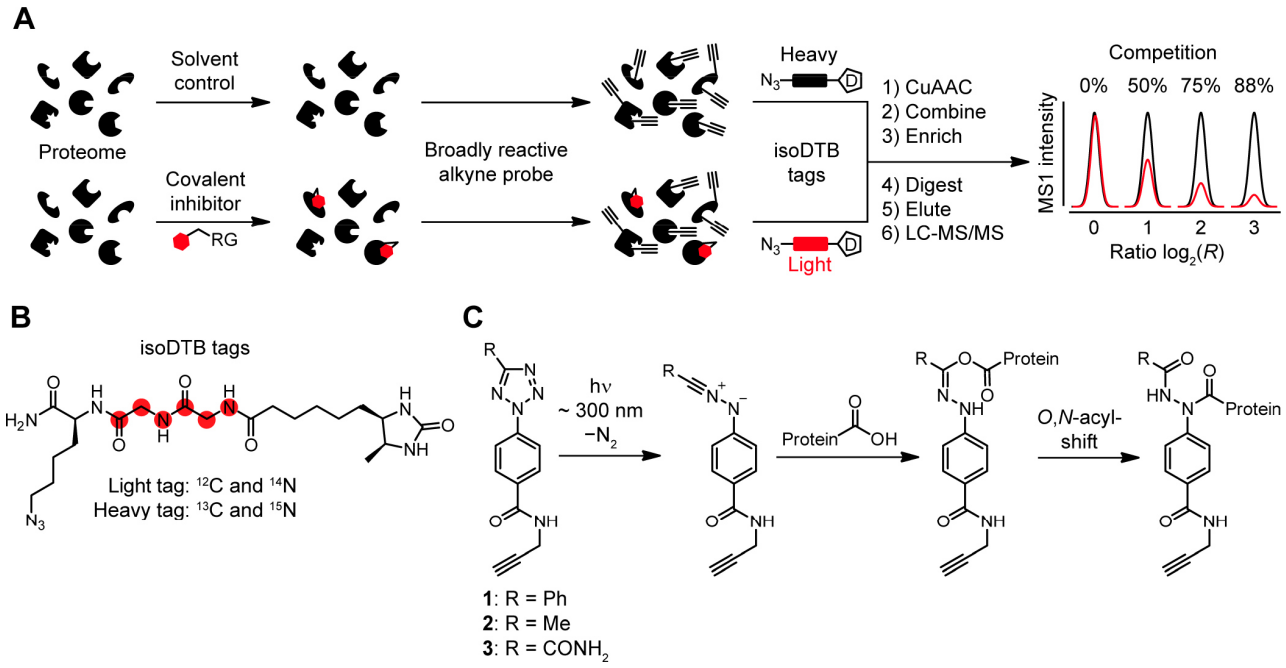


Figure 1. Concept of this study. (A) Workflow of competitive residue-specific proteomics using the isotopically labeled desthiobiotin azide (isoDTB) tags.[11] RG, reactive group; D, desthiobiotin. (B) Structure of the isoDTB tags.[11] (C) Light-induced reactivity of 2,5-disubstituted tetrazoles **1**−**3** with carboxylic acids in proteins. While other nucleophiles might attack the nitrilimine, only for carboxylic acids a stable product can be formed via an *O*,*N*-acyl-shift.

(LC-MS/MS). Because the peptides modified with the probe are directly detected, not only the target protein of the covalent inhibitor but also its exact interaction site are identified. Here, residues that are not bound by the inhibitor will show ratios of around one ($R \approx 1$), whereas residues that are strongly engaged by the covalent inhibitor will show high ratios ($R \gg 1$). In this way, quantitative information on the sites that are modified by the covalent inhibitor is obtained.

In order to expand the scope of the described residue-specific proteomics workflow to amino acid residues other than cysteine, broadly reactive alkyne probes are needed that specifically address a subset of amino acids of interest. For cysteines, iodoacetamide-alkyne (IA-alkyne) is most widely used.[12] For lysines, acylation reagents have been shown to efficiently label a large number of residues in the proteome.[8,13] Recently, sulfonyl triazoles have successfully been applied to study tyrosines.[14] Many other chemistries have been developed to target, e.g., lysines,[8,13,15−17] tyrosines,[18−20] methionines,[21,22] histidines,[23] tryptophanes,[24] as well as aspartates and glutamates[25−27] in proteins, but their exploration as broadly reactive alkyne probes for residue-specific proteomics is still lacking.

We were especially enticed by targeting aspartates and glutamates in the proteome as these amino acids frequently occur in the bacterial proteome (~12% of all residues),[9] tend to be in pockets or on the proteome surface due to their polarity, and could show unique reactivity over all other nucleophilic amino acids as their initially nucleophilic character can be turned into an electrophilic reactivity by suitable activators. Nevertheless, only a few chemotypes including sulfonate esters,[28] diazonium salts,[27] and oxazolium salts[25,26] have been studied for the development of carboxylic-acid-directed covalent inhibitors. Therefore, we reasoned that the development of additional selective chemotypes could be fostered by the availability of an isoDTB-ABPP-based platform that allows investigating aspartates and glutamates proteome-wide.

Previously, isoxazolium salts based on Woodward's reagent K have been investigated as probes for a chemoproteomic study of carboxylic acids.[26] These studies led to the identification of a few target proteins, but residue-specific information could not be obtained on a global level.

Another very interesting chemotype for targeting carboxylic acids is light-activatable 2,5-disubstituted tetrazoles (Figure 1C).[29] These compounds are inherently unreactive but, upon light activation, produce nitrilimines through liberation of nitrogen.[29] These nitrilimines are highly reactive intermediates that were originally described as bio-orthogonal moieties to address alkenes in proteins.[29] Lately, it has been shown that they can also react with a number of proteinogenic nucleophiles including carboxylic acids.[30−33] This prompted us to explore their usage as broadly reactive alkyne probes. These probes would have several advantages over other strategies targeting carboxylic acids. First, by changing the substituent at the 5-position, their electronic properties can be fine-tuned, which should allow adjusting their reactivity, selectivity, and target profile.[31] Second, they would form very stable 1,2-diacyl-1-arylhydrazines after reaction with carboxylic acids, which should facilitate the chemoproteomics analysis.[33] Third, they are unreactive and therefore stable and potentially nontoxic before activation, which could enable their usage in living cells as has been previously shown for light-activatable cysteine-directed electrophiles.[34] 2,5-Disubstituted tetrazoles have been used in a proteomic context before in order to broadly identify their target proteins, but these studies have not investigated the global specificity of the probes toward certain amino acids or looked at the interactions in a residue-specific manner.[33]

Therefore, we set out to study the reactivity and selectivity of 2,5-disubstituted tetrazoles in a proteome-wide context. We

https://dx.doi.org/10.1021/acscentsci.9b01268
ACS Cent. Sci. 2020, 6, 546−554

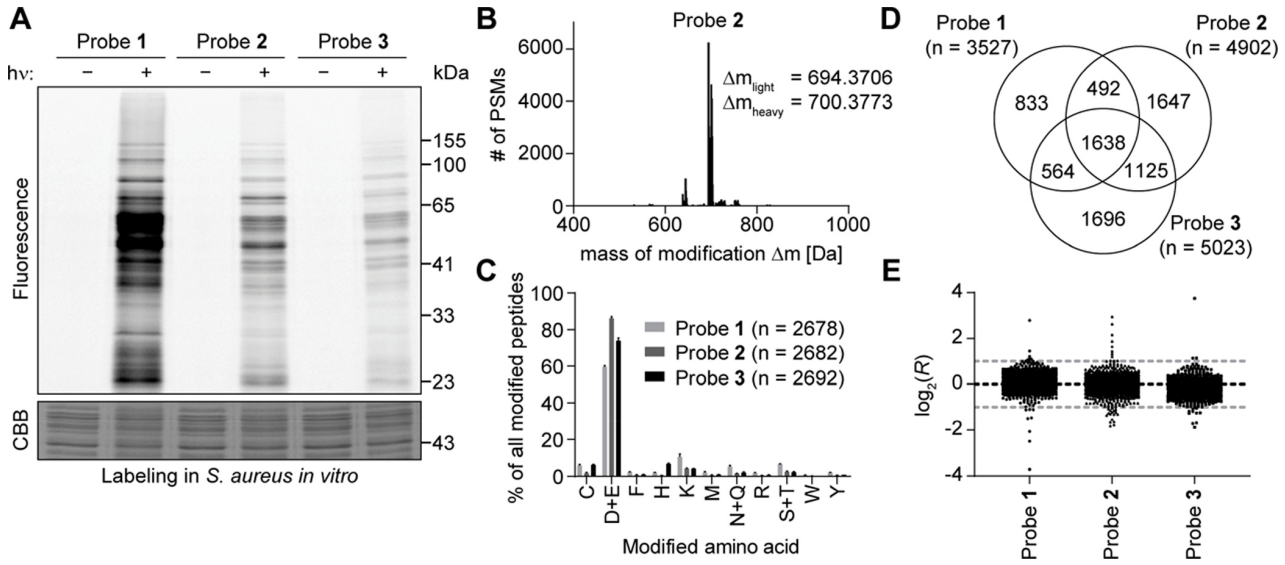


Figure 2. Light-activatable 2,5-disubstituted tetrazoles allow global monitoring of aspartates and glutamates in the *S. aureus* proteome *in vitro* with high specificity. (A) Gel-based analysis of labeling with probes **1**−**3**. *S. aureus* lysate was treated with 100 μM of the indicated probe, incubated for 30 min, irradiated with light (λ = 280−315 nm) for 10 min, and labeled with TAMRA-azide using CuAAC. Controls were performed without irradiation. Gel-based analysis was performed with in-gel fluorescence scanning and staining using Coomassie Brilliant Blue (CBB). (B) Analysis of the mass of modification on tryptic peptides after labeling of *S. aureus* lysate with 100 μM probe **2**. MSFragger software[35] was used to determine, which masses of modification occur in the proteomic samples labeled with probe **2** after light activation and CuAAC to the light and heavy isoDTB tags. Expected masses of modification for tryptic peptides labeled with **2** according to the reactivity shown in Figure 1C and additionally modified with light or heavy isoDTB tag, respectively, are 694.3663 and 700.3738 Da. PSM: peptide-spectrum match. (C) Analysis of the amino acid specificity of the probes. Proteomes labeled with the indicated probe after light activation and modified by CuAAC with the light and heavy isoDTB tags were analyzed with MaxQuant software[36] allowing the modification on any potentially nucleophilic amino acid. Peptides were included in the analysis if the localization probability for a single residue was more than 75%. Data shows the mean ± the standard deviation. The total number of identified PSMs is given in parentheses. (D) Venn diagram of the number of quantified aspartates and glutamates with the three different probes. (E) Plot of the ratios $\log_2(R)$ for aspartates and glutamates in proteomic samples, in which the heavy- and light-labeled sample were both modified with 100 μM of the indicated probe without pretreatment with an inhibitor. The expected value of $\log_2(R) = 0$ is indicated by the black line; the preferred quantification window ($-1 < \log_2(R) < 1$) is indicated by the two gray lines. Each dot represents one quantified aspartate or glutamate. All data for panels B−E originates from biologically independent duplicates of technical duplicates.

demonstrate their ability to act as broadly reactive alkyne probes to study aspartates and glutamates *in vitro* and *in situ* with high specificity even in challenging Gram-negative bacteria. Furthermore, we study the binding of covalent ligands and introduce a new class of carboxylic-acid-directed protein ligands, namely, hydrazonyl chlorides.

RESULTS

Synthesis of 2,5-Disubstituted Tetrazoles. In order to investigate the proteome-wide reactivity of 2,5-disubstituted tetrazoles, we set out to synthesize three different probes (**1**−**3**, Figure 1C). Due to the different effects of the substituents at the 5-position (aromatic phenyl group for **1**, aliphatic methyl group for **2**, and electron-withdrawing carboxamide group for **3**), we reasoned that these probes should allow us to tailor their reactivity and selectivity.[31] All three probes were synthesized according or similar to literature-known procedures (Scheme S1).[33] For probes **1** and **3** we synthesized the diazonium salt starting from *para*-aminobenzoic acid and reacted it with benzaldehyde phenylsulfonylhydrazone or ethyl glyoxylate *para*-tosylhydrazone to give the respective tetrazoles.[33] These were coupled to propargyl amine using *N*-(3-dimethylaminopropyl)-*N*′-ethylcarbodiimide hydrochloride (EDC·HCl) to give probe **1** and the ethyl ester precursor of probe **3**. This precursor was reacted with ammonia to give the final carboxamide **3**. For the synthesis of probe **2**, we synthesized the diazonium salt starting from methyl *para*-amino-benzoate and reacted this with acetamidine hydrochloride using subsequent oxidation with iodine and potassium iodide.[37] Saponification of the ester with sodium hydroxide and coupling to propargyl amine using EDC HCl yielded probe **2**.[33]

We first studied the activation of the probes to the nitrilimine in PBS solution by UV light using LC-MS detection. For all three probes, we detected quantitative light activation followed by hydrolysis (probes **1**−**3**) or reaction with chloride (probe **3**) within 10 min of UV irradiation at 280−315 nm (Figure S1) detection indicating efficient activation.[31]

Aspartate and Glutamate-Specific Proteome-wide Labeling with 2,5-Disubstituted Tetrazoles. With these probes in hand, we next evaluated their proteome-wide reactivity in the *Staphylococcus aureus* (*S. aureus*) strain SH1000. After optimization of the irradiation time (Figure S2) and probe concentration (Figure S3) using a gel-based readout, we directly compared the labeling with the three probes at 100 μM with irradiation for 10 min at 280−315 nm (Figure 2A). For all three probes, we could detect strong labeling of many different proteins in the gel. All probes are therefore in principle suitable to investigate many binding sites in proteins. We observed no striking differences in the overall labeling pattern but stronger labeling with probe **1** in comparison to probe **2** and **3**, which is most likely caused by its increased lipophilicity.

As all probes showed promising labeling using gel-based detection, we set out to investigate their reactivity and selectivity

https://dx.doi.org/10.1021/acscentsci.9b01268
ACS Cent. Sci. 2020, 6, 546−554

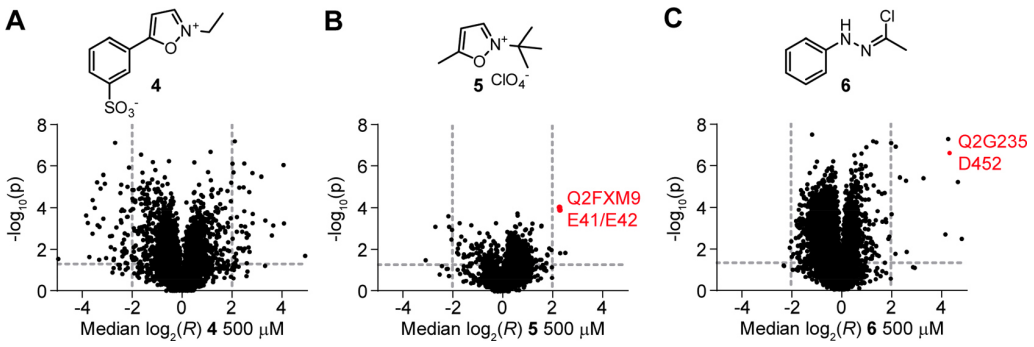


Figure 3. Probe **2** reveals the targeted residues of carboxylic-acid-directed covalent protein ligands proteome-wide using isoDTB-ABPP *in vitro*. (A–C) Volcano plots of isoDTB-ABPP experiments comparing samples pretreated with 500 μM of the indicated covalent ligand to a solvent control. Plotted are the ratio ($\log_2(R)$) between the heavy (solvent-treated, 1% HCl for **4**, DMSO for **5**, DMF for **6**) and light (compound-treated) labeled samples and the probability in a one-sample *t* test that *R* is equal to one ($-\log_{10}(p)$). The targeted E41/E42 of pyruvate kinase (UniProt code: Q2FXM9) and D452 of nicotinate phosphoribosyltransferase (UniProt code: Q2G235) are highlighted in red. All data originates from two (B and C) or three (A) biologically independent experiments; performed in technical duplicates (A and B) or triplicates (C).

in a mass-spectrometry-based setup using the isoDTB-ABPP workflow (Figure 1A). Here, we treated both samples with the same concentration of the probe and did not pretreat with an inhibitor. As the two samples are in this case identical, the isotopic ratios *R* for all residues are expected to be around one, which gives an additional quality control and can be used to investigate if the method gives quantitative results.

First, we wanted to establish in an unbiased way if we detect peptides with the expected modification in the samples. For this purpose, we utilized the MSFragger software that allows an Open Search, in which peptides are identified and assigned their modification state without the need to predefine a modification (Figure 2B, and Figure S4 and Table S1).[35] Indeed, in all cases we were able to mainly detect the modification with the expected mass for the 1,2-diacyl-1-aryl-hydrazine moiety of the respective probe modified with the light or heavy isoDTB tag, respectively. This indicates that the expected reactivity via the nitrilimine intermediate is directly observed using the chemoproteomic setup.

Next, we set out to investigate the amino acid specificity of the probes. For this purpose, we used MaxQuant software[36] and allowed the modification with the respective probe to be on any potentially nucleophilic amino acid (C, D, E, F, H, K, M, N, Q, R, S, T, W, Y). We further analyzed all peptides with a high localization probability for a single residue (>75%, analogous to class I phosphosites) as determined by the Andromeda algorithm implemented in MaxQuant that uses a probabilistic scoring model for the localization.[38] The median localization probability for all analyzed peptides across all probes was >95%. We manually verified a selection of annotated modification sites using the respective MS2 spectra (Figure S5). For all three probes (Figure 2C, Table S1), we detected that most peptides are modified on aspartates and glutamates. Interestingly, probe **2** with a methyl group at the 5-position showed an increased selectivity of more than 85% of all peptides modified at aspartates and glutamates. In order to verify this selectivity, we utilized the MSFragger software[35] that uses a fragment-ion indexing algorithm for the localization of the modifications. Here, we analyzed all spectra, for which the modification was uniquely assigned to a single residue, and detected almost exactly the same amino acid selectivity compared to MaxQuant analysis (Figure S6). Probe **2** is therefore a prime candidate as a selective tool for aspartates and glutamate labeling *in vitro*.

In order to better understand the selectivity of probe **2**, we performed experiments with amino acids protected at the *Nα*-position and at the *α*-carboxylic acid in solution. For this purpose, we treated 100 μM of probe **2** with 5 mM of the respective amino acid and investigated the reactivity after 10 min of UV irradiation using LC-MS. In PBS, we detected only very limited adduct formation and mainly hydrolysis in the presence of glutamate (Figure S7). We reasoned that the lipophilic nature of protein binding pockets increases the reactivity of the nitrilimine. We, therefore, studied the reactivity with increasing amounts of acetonitrile (Figure S7).[30,32] At 75% acetonitrile, adduct formation with glutamate was detected as the main peak. Under these conditions also aspartate mainly reacted to give the expected product (Figure S8). While serine and tyrosine exclusively showed hydrolysis of the nitrilimine, some adduct was formed for lysine and histidine. For cysteine mainly products other than hydrolysis were detected. Here, besides the adduct peak, we detected a peak that had the mass of the thiohydrazide formed from probe **2**. This corresponds to a formal thiolysis of the nitrilimine. We reasoned that, in analogy to the light-induced reactivity of 2-thiazolines,[39] this indicates an initial adduct formation with cysteine followed by homolytic bond cleavage of the adduct. When we performed the reaction with different irradiation times (Figure S9), the activation of probe **2** was quantitative after 7 min. Next to the hydrolysis product, this mainly resulted in the formation of the cysteine adduct and the thiohydrazide. The cysteine adduct was converted to the thiohydrazide upon further irradiation. This process was almost quantitative after 20 min of irradiation. This shows that the instability of the cysteine adduct toward UV irradiation might contribute to the selectivity of probe **2** for aspartate and glutamate labeling. As the final reaction product with cysteine does not lead to a modification with an alkyne, this reactivity does not compromise the use of 2,5-disustituted tetrazoles as probes for proteome-wide monitoring of carboxylic acids. Nevertheless, for the potential development of covalent inhibitors, this reactivity will have to be carefully considered.

Finally, we used MaxQuant software to also quantify the modified aspartates and glutamates.[36] For this purpose, we allowed modification only at these residues. We were able to quantify more than 3500 aspartates and glutamates with each of the three probes (Figure 2D, Table S1). Probes **2** and **3** even quantified around 5000 residues, each. As the heavy and light

https://dx.doi.org/10.1021/acscentsci.9b01268
ACS Cent. Sci. 2020, 6, 546–554

samples were mixed at a ratio of 1:1, all residues are expected to be quantified with an R value close to one ($\log_2(R) \approx 0$). In all cases, the number of peptides that were outside of the preferred window of $-1 < \log_2(R) < 1$ was ≤1% indicating that the probes allow reliable residue-specific quantification (Figure 2E, Table S1). There is a significant overlap between the residues detected with each probe (Figure 2D), but there are also many residues that are exclusively detected with one of the probes. Furthermore, digestion with chymotrypsin instead of trypsin in an additional experiment led to an additional increase in the total number of quantified residues for probe **2** to 6136 (Figure S10 and Table S1). By combining the data from the three probes and the additional experiments with chymotrypsin digestion for probe **2**, we were able to quantify a total of 8971 aspartates and glutamates in the proteome of *S. aureus*. It is striking that probe **1** allowed quantification of fewer aspartates and glutamates, while showing the most intense labeling by gel-based experiments (Figure 2A). Due to the lower specificity of the probe, it can be speculated that this number is lowered by the higher number of peptides modified at other sites as the total amount of detected modified peptides is similar between all probes (Figure 2C). In order to obtain an estimate for the sensitivity of the method, we added bovine serum albumin (BSA) at different concentrations to the *S. aureus* lysate and performed isoDTB-ABPP with probe **2**. We were able to detect BSA down to a concentration of 10 nM (Table S1). In summary, probe **2** is an ideal tool for the global investigation of carboxylic acids in the bacterial proteome *in vitro*.

Residue-Specific Profiling of Covalent Protein Ligands Targeting Aspartates and Glutamates. With probe **2** as an optimized probe in hand, we next wanted to take first steps toward globally investigating carboxylic-acid-directed protein ligands in the proteome of *S. aureus*. For this purpose, we initially investigated established carboxylic-acid-directed chemistry. Here, isoxazolium salts have been used as protein ligands in various instances.[25,26] We decided to investigate Woodward's reagent K (**4**, Figure 3A) and isoxazolium salt **5** (Figure 3B). By performing the chemoproteomic workflow with these ligands at 500 μM (and additionally at 200 μM for **4**, Figure S11 and Table S2) and probe **2** as an optimized broadly reactive alkyne probe, we were able to identify 44 aspartates and glutamates that are able to interact with these compounds in the *S. aureus* proteome. These hits, e.g., include the interaction of **5** with the residues E41/E42 in the ATP-binding site of the essential protein pyruvate kinase (UniProt code, Q2FXM9; Figure 3B, Table S2). For compound **4**, we also detected 50 residues that increase in labeling upon compound treatment. For eight of these residues, we detected another aspartate or glutamate in the same protein that is engaged by at least 50% indicating that this engagement might lead to a structural change causing an increased labeling. Put together, these competitive studies show that our method allows target engagement studies at aspartates and glutamates in the whole proteome and could verify that isoxazolium salts engage these residues also in bacterial proteins.

Enticed by the possibility to address aspartates and glutamates in the proteome with nitrilimines, we next explored if the reactive nitrilimine could also be generated without the use of light and therefore used as a covalent protein ligand. Previous studies that used nitrilimines for conjugation to alkenes have utilized hydrazonyl chlorides for this purpose, but to the best of our knowledge these chemotypes have not been used as carboxylic-acid-directed protein ligands.[40] We therefore synthesized compound **6** (Scheme S2), which can liberate a nitrilimine through elimination of hydrochloric acid. Using **6** at 500 μM in competitive experiments with probe **2**, we identified 13 residues that interact with **6** in the *S. aureus* proteome (Figure 3C, Table S2). These peptides include the residue D452 of the essential protein nicotinate phosphoribosyl-transferase (Npt) (UniProt code: Q2G235). In order to prove this interaction, we recombinantly expressed and purified this protein. Upon treatment with **6** *in vitro*, we were able to detect quantitative modification of Npt and a shift in the mass of the protein that corresponds to the modification with one molecule of the nitrilimine formed from **6** (Figure S12). To a lower degree, we were also able to detect double and triple modification. As the quite reactive hydrazonyl chloride was used at a high concentration on an isolated protein, this points to the fact that multiple carboxylic acid residues can be modified by **6** in this setup.

In order to verify the binding sites on Npt and quantify their engagement, we used a recently established protocol based on reductive dimethyl labeling for quantification.[8] After treatment of recombinant Npt with **6**, we were able to detect the **6**-modified peptide with modification at the expected residue D452 and quantified the target engagement to approximately 50% (Figure S13 and Table S2), which is the highest engagement among all quantified aspartates and glutamates in Npt. Nevertheless, we were also able to detect modified peptides at other sites, which points to the fact that other residues are also reactive when compound **6** is used on an isolated protein at this high concentration.

Despite the considerable target engagement at D452, we could detect no effect on enzyme activity (Figure S14), pointing toward the fact that Npt is not inhibited by covalent modification of this residue by **6**. Nevertheless, this experiment shows that hydrazonyl chlorides are interesting warheads for targeting carboxylic acids in proteomes and that the product of the reaction with carboxylic acids in proteins is a 1,2-diacyl-1-arylhydrazine, which is in agreement with the formation of an intermediate nitrilimine.

Through compiling of the proteome-wide isoDTB-ABPP data for all three competitors (**4**–**6**), we detected a total of 56 aspartates and glutamates in 48 different proteins that are targeted by at least one of these compounds in the proteome of *S. aureus*. These residues show a slightly increased probability to be at functional sites (5.4%) compared to other quantified residues (4.0%) or the genomic background (2.7%, Figure S15). This suggests that in functional sites higher residue reactivity or additional noncovalent interactions might facilitate high occupancy binding of the ligands. Due to the limited number of identified liganded residues with the used three competitors, this observation will need to be verified once a larger data set of ligandable residues is available. Interestingly, residues in essential proteins are strongly enriched among the quantified residues compared to the genomic background (Figure S16). Concerning the functional classes of proteins, the liganded residues and other quantified residues show a very similar distribution to each other, in which enzymes and proteins involved in gene expression are enriched, and receptors, transporters, and channels are depleted in comparison to the whole genome (Figure S17). These facts point to the chance to functionally target diverse important proteins with these compounds. Comparing the specific interactions of the three compounds, one can see that most residues are exclusively targeted by one of the compounds indicating some specificity even with these very small compounds at high concentrations

https://dx.doi.org/10.1021/acscentsci.9b01268
ACS Cent. Sci. 2020, 6, 546−554

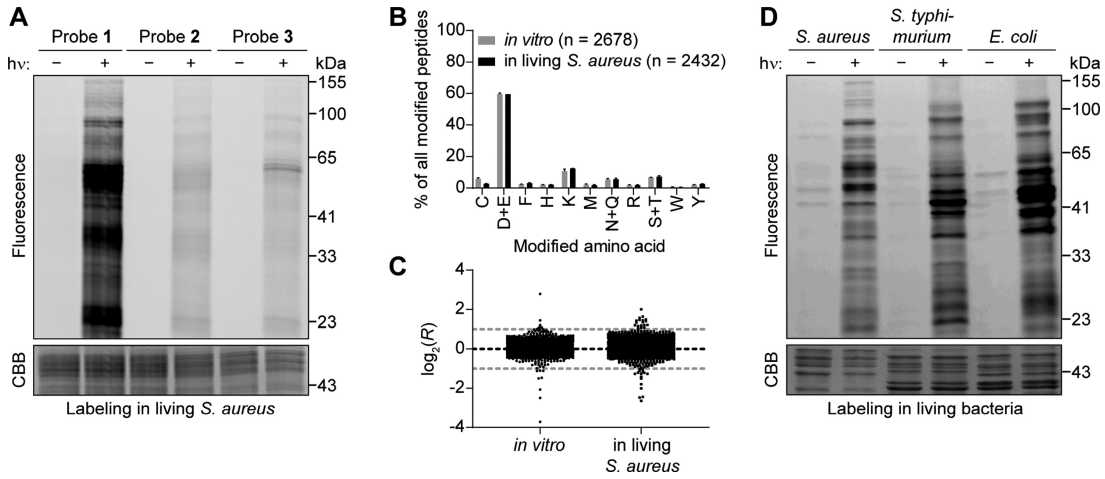


Figure 4. Specific labeling of aspartates and glutamates with 2,5-disubstituted tetrazoles in living bacteria. (A) Probe **1** efficiently labels proteins in living *S. aureus*. Living bacteria were treated with the indicated probe for 1 h and labeled by irradiation with light (λ = 280−315 nm) for 10 min. After lysis, TAMRA azide was attached using CuAAC, and labeling was analyzed using gel electrophoresis with in-gel fluorescence scanning and Coomassie Brilliant Blue (CBB) staining. Controls were performed without irradiation. (B) Analysis of the amino acid specificity of probe **1** *in vitro* and in living *S. aureus*. Samples were analyzed with MaxQuant software[36] allowing the modification with probe **1** and the heavy or light isoDTB tag to be on any potentially nucleophilic amino acid. Peptides were further analyzed if the localization probability for a single residue was more than 75%. The total number of identified modification sites is given in parentheses. Data shows the mean ± the standard deviation. (C) Plot of the ratios $\log_2(R)$ of samples, in which the heavy- and light-labeled sample were both modified with the same concentration of the indicated probe in living *S. aureus* without pretreatment with an inhibitor. The expected value of $\log_2(R) = 0$ is indicated by the black line; the preferred quantification window ($-1 < \log_2(R) < 1$) is indicated by the two gray lines. One data point in living bacteria at $\log_2(R) = -8.9$ is not shown for clarity. All data for panels B and C originates from biologically independent duplicates of technical triplicates for data in living bacteria and from biologically independent duplicates of technical duplicates for data *in vitro*. (D) Probe **1** efficiently labels proteins in the living Gram-negative bacteria *S. typhimurium* and *E. coli*. The experiment was performed in the indicated bacteria as described for *S. aureus* in part A.

(Figure S18). Chemoproteomic profiling of proteomes with probe **2** is therefore a very promising strategy for the optimization of carboxylic-acid-directed protein ligands.

Global Investigation of Aspartates and Glutamates in Living Bacteria. Having a working platform in hand that allows quantifying many aspartates and glutamates *in vitro*, we next wanted to investigate if this technology can also be transferred to monitor these amino acids in living bacteria. As the probes are nonreactive before light activation, we reasoned that they could be nontoxic in bacteria. We therefore measured the minimum inhibitory concentration (MIC) for all three probes for the growth of *S. aureus* (Figure S19). All three compounds had no MIC or effect on the optical density of the culture up to a concentration of 200 μM. The probes can therefore be used in living bacteria without affecting their viability.

We next checked for labeling *in situ* using a gel-based readout. Here, we pretreated living *S. aureus* cells with the probes and then labeled the aspartates and glutamates using UV irradiation. After optimization of the pretreatment (Figure S20) and irradiation time (Figure S21) using probe **1**, we directly compared all three probes (Figure 4A). We could detect that there is only very weak labeling with probes **2** and **3** in comparison to probe **1**. The difference is much larger than *in vitro* (Figure 2A). Probe **1** therefore seems to be taken up into cells much more efficiently than the other probes. We, therefore, decided to utilize probe **1** for *in situ* experiments, although the selectivity for aspartates and glutamates is not quite as high, because of the observed striking difference in labeling intensity.

Next, we performed a mass-spectrometry-based experiment as described above, in which two samples of living bacteria were individually treated with probe **1**, irradiated, lysed, clicked to the light and heavy isoDTB tags, respectively, and mixed at a ratio of 1:1 before analysis. We could verify the same selectivity for aspartates and glutamates as seen *in vitro* for probe **1** (Figure 4B, Table S1). We were able to quantify 3928 residues with the number of incorrectly quantified peptides again being ≤1% (Figure 4C, Table S1). The method is therefore applicable to also quantitatively monitor aspartates and glutamates in living *S. aureus*.

Encouraged by this result, we also investigated the Gram-negative bacteria *Escherichia coli* (*E. coli*) and *Salmonella typhimurium* (*S. typhimurium*). Gram-negative bacteria possess a very strong outer barrier consisting of the cell wall and two cell membranes as well as efficient efflux mechanisms. This renders the development of inhibitors and probes that efficiently engage proteins in these cells very challenging.[6] After testing the three probes by gel-based experiments and demonstrating that probe **1** again gives the strongest labeling (Figures S22 and S23), we directly compared the labeling in living *S. aureus*, *E. coli*, and *S. typhimurium* using probe **1** (Figure 4D). Probe **1** labeled the proteomes of the Gram-negative bacteria to a similar degree as in *S. aureus*, which indicates that this probe is efficiently taken up also into these most challenging cells. Performing a mass-spectrometry-based experiment in *E. coli*, we were able to quantify 1905 residues with good selectivity (Figure S24 and Table S1). Our platform therefore allows global profiling of aspartates and glutamates also in challenging Gram-negative bacteria, which will have important implications for drug discovery of covalent inhibitors through target engagement studies.

■ DISCUSSION

We here describe the first method to globally map aspartates and glutamates in a residue-specific fashion. We synthesized three

https://dx.doi.org/10.1021/acscentsci.9b01268
ACS Cent. Sci. 2020, 6, 546−554

light-activatable 2,5-disubstituted tetrazole probes and investigated them as broadly reactive alkyne probes to study the proteome-wide interactions of covalent, carboxylic-acid-directed protein ligands using isoDTB-ABPP.

Initially, we optimized the labeling conditions *in vitro* in the proteome of *S. aureus* and could show that probe **2** exhibits the highest selectivity with >85% of all peptides being labeled at aspartates and glutamates. In this way, we were able to quantify more than 6000 residues using this probe and, in aggregate, more than 8500 residues across all probes. This corresponds to 8.9% of all aspartates and glutamates encoded in the *S. aureus* genome. Furthermore, we quantify 398 residues at functional sites including 174 residues at functional sites of essential proteins. Residues at functional sites of essential proteins (Figure S5) include, e.g., E293 at the ATP binding site of phosphoglycerate kinase (UniProt code: Q2G031), E179 at the active site of the GMP synthase GuaA (UniProt code: Q2G0Y6), E197 at a magnesium binding site of succinate-CoA ligase (UniProt code: Q2FZ37), D233 at the active site of FabH (UniProt code: Q2FZS0), E330 at the ATP binding site of the glycine-tRNA ligase GlyQS (UniProt code: Q2FY08), and E453 at the substrate binding site of 6-phosphogluconate dehydrogenase (UniProt code: Q2FY60). In this way, the method bears great potential to identify functionally important effects on carboxylic acids in the bacterial proteome. Future studies should be able to further increase this coverage of carboxylic acids and especially functional sites. For this purpose, the proteome can be prefractionated using, e.g., strong cation exchange or high pH fractionation in order to decrease the complexity of the samples in individual MS experiments. Furthermore, it is tempting to speculate that more chemically diverse 2,5-disubstituted tetrazoles can be synthesized to cover more functional sites by specific interactions. The 2,5-disubstituted tetrazole probe **2** offers great specificity and high coverage for aspartate and glutamate residues in lysates. Thus, it allows for the first time a broad study of effects on aspartates and glutamates in the proteome.

Next, we used this technology to map the interactions of carboxylic-acid-directed protein ligands. Here, we first profiled two isoxazolium salts (**4** and **5**) and identified several residues that are modified by these compounds. These include many residues at functional sites and in essential proteins. Furthermore, we proposed, synthesized, and evaluated hydrazonyl chlorides as a new class of carboxylic-acid-directed protein ligands. This design was based on the fact that they can liberate nitrilimines upon elimination of hydrochloric acid that are similar to those resulting from our light-activatable probes. We synthesized one member of this compound class (**6**) and found 13 targeted binding sites in the proteome of *S. aureus*. We verified the interaction of **6** with the recombinant, essential protein nicotinate phosphoribosyltransferase *in vitro* and verified the expected binding site and mass of the modification on the protein. Nevertheless, due to the high reactivity and concentration of **6**, some additional labeling of other sites was also observed indicating that further optimization of this chemotype will be necessary. Taken together, this shows that our method gives residue-specific target engagement information for carboxylic-acid-directed protein ligands in the bacterial proteome.

So far, all three tested competitors did not show antibacterial activity up to 500 μM (Figure S25). Compound **6** is nevertheless a most promising candidate for further optimization into a specific carboxylic-acid-directed covalent inhibitor with biological activity as it shows specific engagement of several residues. Besides varying the substituents of the molecule in order to introduce binding to specific proteins and to tailor its reactivity, it would be especially interesting to investigate various leaving groups instead of chloride in order to further tune the stability, reactivity, and selectivity of these compounds as has recently been shown for targeting tyrosines.[14] Furthermore, future development will have to carefully investigate the possibility to avoid reactivity with cysteine residues in order to develop inhibitors that can be used in biologically relevant settings. The presented method will be instrumental in evaluating these compounds in a proteome-wide sense and using them to generate a more complete map of aspartates and glutamates that can be addressed with covalent ligands.

Finally, we could show that probe **1** can be used to monitor close to 4000 aspartates and glutamates with good selectivity when it is used in living *S. aureus* cells. The probe even allows obtaining a broad map of these residues in living Gram-negative bacteria, which are notoriously hard to penetrate with chemical probes. In this way, the methodology will allow obtaining important insights into the behavior of aspartates and glutamates in this biologically relevant setting.

Taken all of this together, we present a method that for the first time allows to globally and residue-specifically monitor aspartates and glutamates in the bacterial proteome *in vitro* and in living bacteria. The method furthermore allows the identification of sites that can be addressed with covalent ligands in the proteome. In this way, we are convinced that this methodology will have important implications for the design of new chemotypes for covalent inhibitors that target carboxylic acids as well as to broadly understand the target engagement at these amino acid residues in order to more efficiently develop new antibiotics with a covalent mechanism-of-action.

While this manuscript was under consideration, 2*H*-azirines were reported as additional probes to residue-specifically monitor aspartates and glutamates in the proteome.[41] We are convinced that with these two complementary probe technologies using constitutive and light-activatable electrophiles, respectively, important progress will be made in the design of specific carboxylic-acid-directed covalent inhibitors.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acscentsci.9b01268.

Additional text, figures, schemes, and tables presenting chemoproteomic data, biochemical results, synthetic procedures, experimental conditions, and ^{1}H and ^{13}C NMR spectra (PDF)

Table S1: isoDTB-ABPP data for the evaluation of the mass of modification, selectivity, quantification, and sensitivity for probes **1**−**3** *in vitro* in *S. aureus*, in living *S. aureus*, and in living *E. coli*. (XLSX)

Table S2: Competitive isoDTB-ABPP data for the competitors **4**−**6** *in vitro* in *S. aureus* as well as of reductive dimethyl labeling experiments to establish and quantify the site of labeling on recombinant nicotinate phosphoribosyltransferase (XLSX)

Table S8: Annotation database for all aspartates and glutamates in the genome of *S. aureus* SH1000 (XLSX)

https://dx.doi.org/10.1021/acscentsci.9b01268
ACS Cent. Sci. 2020, 6, 546−554

AUTHOR INFORMATION

Corresponding Author

Stephan M. Hacker − *Department of Chemistry, Technical University of Munich, Garching D-85747, Germany;* orcid.org/0000-0001-5420-4824; Email: stephan.m.hacker@tum.de

Authors

Kathrin Bach − *Department of Chemistry, Technical University of Munich, Garching D-85747, Germany;* orcid.org/0000-0003-0284-4218

Bert L. H. Beerkens − *Department of Chemistry, Technical University of Munich, Garching D-85747, Germany;* orcid.org/0000-0002-2884-3961

Patrick R. A. Zanon − *Department of Chemistry, Technical University of Munich, Garching D-85747, Germany;* orcid.org/0000-0002-8883-8275

Complete contact information is available at: https://pubs.acs.org/10.1021/acscentsci.9b01268

Author Contributions

[§]K.B., B.L.H.B., and P.R.A.Z. contributed equally to this work. K.B. and B.L.H.B. performed MS- and gel-based experiments. K.B. cloned and expressed nicotinate phosphoribosyltransferase and performed biochemical experiments. B.L.H.B. and P.R.A.Z. synthesized compounds. P.R.A.Z. performed reactivity assays with amino acids using LC-MS. K.B., B.L.H.B., and S.M.H. acquired and analyzed MS data. S.M.H. conceived and supervised the project. S.M.H. wrote the manuscript with contribution from all authors.

Funding

S.M.H. and P.R.A.Z. acknowledge funding by the Fonds der Chemischen Industrie through a Liebig Fellowship and a Ph.D. fellowship. We acknowledge financial support by the TUM Junior Fellow Fund.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We gratefully acknowledge Prof. Dr. Stephan A. Sieber and his group for their generous support.

ABBREVIATIONS

ABPP, activity-based protein profiling; CBB, Coomassie Brilliant Blue; CuAAC, copper-catalyzed azide−alkyne cycloaddition; DMF, dimethylformamide; DMSO, dimethyl sulfoxide; *E. coli*, *Escherichia coli*; EDC·HCl, *N*-(3-dimethylaminopropyl)-*N*′-ethylcarbodiimide hydrochloride; HCl, hydrochloric acid; IA-alkyne, iodoacetamide alkyne; isoDTB tags, isotopically labeled desthiobiotin azide tags; isoDTB-ABPP, isotopically labeled desthiobiotin azide tag-based activity-based protein profiling; isoTOP-ABPP, isotopic tandem orthogonal proteolysis activity-based protein profiling; LC-MS/MS, liquid chromatography coupled to tandem mass spectrometry; MIC, minimum inhibitory concentration; MS, mass spectrometry; PSM, peptide-spectrum match; *R*, ratio between the heavy and the light channel in isoDTB-ABPP experiments; *S. aureus*, *Staphylococcus aureus*; *S. typhimurium*, *Salmonella typhimurium*, TEV, tobacco-etch virus

REFERENCES

(1) Singh, J.; Petter, R. C.; Baillie, T. A.; Whitty, A. The resurgence of covalent drugs. *Nat. Rev. Drug Discovery* **2011**, *10*, 307−317.

(2) Chaikuad, A.; Koch, P.; Laufer, S. A.; Knapp, S. The Cysteinome of Protein Kinases as a Target in Drug Development. *Angew. Chem., Int. Ed.* **2018**, *57*, 4372−4385.

(3) Hendlin, D.; Stapley, E. O.; Jackson, M.; Wallick, H.; Miller, A. K.; Wolf, F. J.; Miller, T. W.; Chaiet, L.; Kahan, F. M.; Foltz, E. L.; Woodruff, H. B.; Mata, J. M.; Hernandez, S.; Mochales, S. Phosphonomycin, a new antibiotic produced by strains of streptomyces. *Science* **1969**, *166*, 122−123.

(4) Nishimura, H.; Mayama, M.; Komatsu, Y.; Kato, H.; Shimaoka, N.; Tanaka, Y. Showdomycin, a New Antibiotic from a Streptomyces Sp. *J. Antibiot.* **1964**, *17*, 148−155.

(5) Smith, P. A.; Koehler, M. F. T.; Girgis, H. S.; Yan, D.; Chen, Y.; Chen, Y.; Crawford, J. J.; Durk, M. R.; Higuchi, R. I.; Kang, J.; Murray, J.; Paraselli, P.; Park, S.; Phung, W.; Quinn, J. G.; Roberts, T. C.; Rouge, L.; Schwarz, J. B.; Skippington, E.; Wai, J.; Xu, M.; Yu, Z.; Zhang, H.; Tan, M. W.; Heise, C. E. Optimized arylomycins are a new class of Gram-negative antibiotics. *Nature* **2018**, *561*, 189−194.

(6) Lakemeyer, M.; Zhao, W.; Mandl, F. A.; Hammann, P.; Sieber, S. A. Thinking Outside the Box-Novel Antibacterials To Tackle the Resistance Crisis. *Angew. Chem., Int. Ed.* **2018**, *57*, 14440−14475.

(7) Backus, K. M.; Correia, B. E.; Lum, K. M.; Forli, S.; Horning, B. D.; Gonzalez-Paez, G. E.; Chatterjee, S.; Lanning, B. R.; Teijaro, J. R.; Olson, A. J.; Wolan, D. W.; Cravatt, B. F. Proteome-wide covalent ligand discovery in native biological systems. *Nature* **2016**, *534*, 570−574.

(8) Hacker, S. M.; Backus, K. M.; Lazear, M. R.; Forli, S.; Correia, B. E.; Cravatt, B. F. Global profiling of lysine reactivity and ligandability in the human proteome. *Nat. Chem.* **2017**, *9*, 1181−1190.

(9) The UniProt Consortium. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* **2019**, *47*, D506−D515.

(10) Weerapana, E.; Wang, C.; Simon, G. M.; Richter, F.; Khare, S.; Dillon, M. B.; Bachovchin, D. A.; Mowen, K.; Baker, D.; Cravatt, B. F. Quantitative reactivity profiling predicts functional cysteines in proteomes. *Nature* **2010**, *468*, 790−795.

(11) Zanon, P. R. A.; Lewald, L.; Hacker, S. M. Isotopically Labeled Desthiobiotin Azide (isoDTB) Tags Enable Global Profiling of the Bacterial Cysteinome. *Angew. Chem., Int. Ed.* **2020**, *59*, 2829−2836.

(12) Speers, A. E.; Cravatt, B. F. A tandem orthogonal proteolysis strategy for high-content chemical proteomics. *J. Am. Chem. Soc.* **2005**, *127*, 10018−10019.

(13) Ward, C. C.; Kleinman, J. I.; Nomura, D. K. NHS-Esters As Versatile Reactivity-Based Probes for Mapping Proteome-Wide Ligandable Hotspots. *ACS Chem. Biol.* **2017**, *12*, 1478−1483.

(14) Hahm, H. S.; Toroitich, E. K.; Borne, A. L.; Brulet, J. W.; Libby, A. H.; Yuan, K.; Ware, T. B.; McCloud, R. L.; Ciancone, A. M.; Hsu, K.-L. Global targeting of functional tyrosines using sulfur-triazole exchange chemistry. *Nat. Chem. Biol.* **2020**, *16*, 150−159.

(15) Hansen, B. K.; Loveridge, C. J.; Thyssen, S.; Wormer, G. J.; Nielsen, A. D.; Palmfeldt, J.; Johannsen, M.; Poulsen, T. B. STEFs: Activated Vinylogous Protein-Reactive Electrophiles. *Angew. Chem., Int. Ed.* **2019**, *58*, 3533−3537.

(16) Tamura, T.; Ueda, T.; Goto, T.; Tsukidate, T.; Shapira, Y.; Nishikawa, Y.; Fujisawa, A.; Hamachi, I. Rapid labelling and covalent inhibition of intracellular native proteins using ligand-directed N-acyl-N-alkyl sulfonamide. *Nat. Commun.* **2018**, *9*, 1870.

(17) Akcay, G.; Belmonte, M. A.; Aquila, B.; Chuaqui, C.; Hird, A. W.; Lamb, M. L.; Rawlins, P. B.; Su, N.; Tentarelli, S.; Grimster, N. P.; Su, Q. Inhibition of Mcl-1 through covalent modification of a noncatalytic lysine side chain. *Nat. Chem. Biol.* **2016**, *12*, 931−936.

(18) Ban, H.; Gavrilyuk, J.; Barbas, C. F. Tyrosine bioconjugation through aqueous ene-type reactions: a click-like reaction for tyrosine. *J. Am. Chem. Soc.* **2010**, *132*, 1523−1525.

(19) Zhang, J.; Ma, D.; Du, D.; Xi, Z.; Yi, L. An efficient reagent for covalent introduction of alkynes into proteins. *Org. Biomol. Chem.* **2014**, *12*, 9528−9531.

(20) Narayanan, A.; Jones, L. H. Sulfonyl fluorides as privileged warheads in chemical biology. *Chem. Sci.* **2015**, *6*, 2650−2659.
(21) Lin, S.; Yang, X.; Jia, S.; Weeks, A. M.; Hornsby, M.; Lee, P. S.; Nichiporuk, R. V.; Iavarone, A. T.; Wells, J. A.; Toste, F. D.; Chang, C. J. Redox-based reagents for chemoselective methionine bioconjugation. *Science* **2017**, *355*, 597−602.
(22) Taylor, M. T.; Nelson, J. E.; Suero, M. G.; Gaunt, M. J. A protein functionalization platform based on selective reactions at methionine residues. *Nature* **2018**, *562*, 563−568.
(23) Jia, S.; He, D.; Chang, C. J. Bioinspired Thiophosphorodichloridate Reagents for Chemoselective Histidine Bioconjugation. *J. Am. Chem. Soc.* **2019**, *141*, 7294−7301.
(24) Seki, Y.; Ishiyama, T.; Sasaki, D.; Abe, J.; Sohma, Y.; Oisaki, K.; Kanai, M. Transition Metal-Free Tryptophan-Selective Bioconjugation of Proteins. *J. Am. Chem. Soc.* **2016**, *138*, 10798−10801.
(25) Martin-Gago, P.; Fansa, E. K.; Winzker, M.; Murarka, S.; Janning, P.; Schultz-Fademrecht, C.; Baumann, M.; Wittinghofer, A.; Waldmann, H. Covalent Protein Labeling at Glutamic Acids. *Cell Chem. Biol.* **2017**, *24*, 589−597.
(26) Qian, Y.; Schurmann, M.; Janning, P.; Hedberg, C.; Waldmann, H. Activity-Based Proteome Profiling Probes Based on Woodward's Reagent K with Distinct Target Selectivity. *Angew. Chem., Int. Ed.* **2016**, *55*, 7766−7771.
(27) Mix, K. A.; Raines, R. T. Optimized diazo scaffold for protein esterification. *Org. Lett.* **2015**, *17*, 2358−2361.
(28) Weerapana, E.; Simon, G. M.; Cravatt, B. F. Disparate proteome reactivity profiles of carbon electrophiles. *Nat. Chem. Biol.* **2008**, *4*, 405−407.
(29) Herner, A.; Lin, Q. Photo-Triggered Click Chemistry for Biological Applications. *Top. Curr. Chem.* **2016**, *374*, 1.
(30) Zhao, S.; Dai, J.; Hu, M.; Liu, C.; Meng, R.; Liu, X.; Wang, C.; Luo, T. Photo-induced coupling reactions of tetrazoles with carboxylic acids in aqueous solution: application in protein labelling. *Chem. Commun.* **2016**, *52*, 4702−4705.
(31) Herner, A.; Marjanovic, J.; Lewandowski, T. M.; Marin, V.; Patterson, M.; Miesbauer, L.; Ready, D.; Williams, J.; Vasudevan, A.; Lin, Q. 2-Aryl-5-carboxytetrazole as a New Photoaffinity Label for Drug Target Identification. *J. Am. Chem. Soc.* **2016**, *138*, 14609−14615.
(32) Li, Z.; Qian, L.; Li, L.; Bernhammer, J. C.; Huynh, H. V.; Lee, J. S.; Yao, S. Q. Tetrazole Photoclick Chemistry: Reinvestigating Its Suitability as a Bioorthogonal Reaction and Potential Applications. *Angew. Chem., Int. Ed.* **2016**, *55*, 2002−2006.
(33) Cheng, K.; Lee, J. S.; Hao, P.; Yao, S. Q.; Ding, K.; Li, Z. Tetrazole-Based Probes for Integrated Phenotypic Screening, Affinity-Based Proteome Profiling, and Sensitive Detection of a Cancer Biomarker. *Angew. Chem., Int. Ed.* **2017**, *56*, 15044−15048.
(34) Abo, M.; Weerapana, E. A Caged Electrophilic Probe for Global Analysis of Cysteine Reactivity in Living Cells. *J. Am. Chem. Soc.* **2015**, *137*, 7087−7090.
(35) Kong, A. T.; Leprevost, F. V.; Avtonomov, D. M.; Mellacheruvu, D.; Nesvizhskii, A. I. MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. *Nat. Methods* **2017**, *14*, 513−520.
(36) Cox, J.; Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **2008**, *26*, 1367−1372.
(37) Ramanathan, M.; Wang, Y. H.; Liu, S. T. One-Pot Reactions for Synthesis of 2,5-Substituted Tetrazoles from Aryldiazonium Salts and Amidines. *Org. Lett.* **2015**, *17*, 5886−5889.
(38) Olsen, J. V.; Blagoev, B.; Gnad, F.; Macek, B.; Kumar, C.; Mortensen, P.; Mann, M. Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* **2006**, *127*, 635−648.
(39) Matsuura, T.; Ito, Y. Photoinduced reactions—LXXIX: Photochemistry of 2-thiazolines. *Tetrahedron* **1975**, *31*, 1245−1250.
(40) Zengeya, T. T.; Garlick, J. M.; Kulkarni, R. A.; Miley, M.; Roberts, A. M.; Yang, Y.; Crooks, D. R.; Sourbier, C.; Linehan, W. M.; Meier, J. L. Co-opting a Bioorthogonal Reaction for Oncometabolite Detection. *J. Am. Chem. Soc.* **2016**, *138*, 15813−15816.
(41) Ma, N.; Hu, J.; Zhang, Z.-M.; Liu, W.; Huang, M.; Fan, Y.; Yin, X.; Wang, J.; Ding, K.; Ye, W.; Li, Z. 2*H*-Azirine-Based Reagents for Chemoselective Bioconjugation at Carboxyl Residues Inside Live Cells. *J. Am. Chem. Soc.* **2020**, *142*, 6051−6059.

https://dx.doi.org/10.1021/acscentsci.9b01268
ACS Cent. Sci. 2020, 6, 546−554

3. Profiling the Proteome-Wide Selectivity of Diverse Electrophiles

Posted as a preprint on *ChemRxiv* and pending peer review.

By Patrick R. A. Zanon, Fengchao Yu, Patricia Z. Musacchio, Lisa Lewald, Michael Zollo, Kristina Krauskopf, Dario Mrdović, Patrick Raunft, Thomas E. Maher, Marko Cigler, Christopher J. Chang, Kathrin Lang, F. Dean Toste, Alexey I. Nesvizhskii, Stephan M. Hacker

https://doi.org/10.26434/chemrxiv.14186561

Synopsis

Even though several broadly reactive alkyne probes that allow monitoring different amino acids proteome-wide have been reported, differences in the experimental workflow, instrumentation, and data analysis hinder direct comparison of reactivity and selectivity. Furthermore, for many of the proteinogenic nucleophilic side chains, such probes are still lacking. To address these issues, we combined our recently developed isoDTB-ABPP platform with the FragPipe software suite into a universal and unbiased workflow that identifies the modification and the protein sites it occurs on in a quantitative fashion.

We benchmarked the new and optimized data analysis features with a published data set of **IA-alkyne** in the lysate of *S. aureus* and confirmed the high selectivity for cysteines and the accurate quantification of these residues. Many additional chemistries to address cysteines exist, such as nucleophilic aromatic substitution, hypervalent iodine reagents, and *Michael* acceptors. All respective probes showed a moderate to high preference for cysteines, with varying degrees of reactivity.

We verified the high selectivity of acylating agents for lysines and described alkyl squaric acid monoamides (**AlkSq-alkyne**) as promising structures for the design of TCIs. Leveraging a cascade reaction of imine formation and cyclization, the ethynylbenzaldehyde **EBA-alkyne** also allows the monitoring of lysines. We additionally identified the photoprobe ***o*NBA-alkyne** as a promising tool to study these residues. While the formylpyridine **PCA-alkyne** showed high selectivity for *N*-termini, the low coverage warrants further optimization.

Building on our previous work on aspartate- and glutamate-targeting probes, we have developed a hydrazonoyl chloride probe (**HC-alkyne**), which performs almost identical to its photoactivatable tetrazole analog (**MeTet-alkyne**). Together with the azirine **Az-alkyne**, this probe will serve as a complementary tool to monitor these residues proteome-wide.

We confirmed the reported selectivities for substitution at sulfur(VI)-centers by the phenolic hydroxyl group of tyrosines. Through electrophilic aromatic substitution, **PTAD-alkyne** modifies tyrosines with high selectivity, but fragmentation of the probe to an isocyanate also leads to modification of lysines and *N*-termini. In the complex environment of bacterial lysates, diazonium salts do not selectively engage tyrosines through the expected diazo coupling. Instead, we detected arylation of cysteines and all aromatic amino acids as the main modification. This surprising finding underlines the power of our unbiased data analysis approach.

Furthermore, we have developed an oxaziridine probe (**OxMet2-alkyne**) with optimized stability of the modification on methionines to enable the monitoring of those residues. The photoactivated *N*-carbamoylpyridinium salt **CP-alkyne** is almost exclusively labeling histidines and tryptophans and is the first probe that allows profiling both those residues proteome-wide. Additionally, *o*-quinone methides generated *in situ* through UV-irradiation also preferably engage tryptophan residues through a formal [4+2]-cycloaddition. Finally, the cyclization of arginines with glyoxals (**PhGO-alkyne**) facilitates the proteome-wide profiling of this amino acid for the first time.

Overall, we synthesized and analyzed over 50 alkyne probes and have identified a set of tools that allow selective monitoring of nine different amino acids and the *N*-terminus proteome-wide. In addition to the experiments in *S. aureus*, we also verified that these probes, except for **PCA-alkyne**, retain their selectivities in the lysate of a human cancer cell line. We are convinced that this selection of probes for competitive residue-specific proteomics will guide the development of new TCIs that engage the target space of protein binding sites lacking suitable cysteine residues.

Author Contributions

Patrick R. A. Zanon and Stephan M. Hacker designed the research, planned experiments, performed and analyzed proteomics experiments. Fengchao Yu and Alexey I. Nesvizhskii designed, developed, and benchmarked data analysis software. Patrick R. A. Zanon synthesized the majority of probes. Patricia Z. Musacchio, Lisa Lewald, Michael Zollo, Kristina Krauskopf, Dario Mrdović, Patrick Raunft, Thomas E. Maher, and Marko Cigler contributed the synthesis of individual probes. Kathrin Lang, Christopher J. Chang, and F. Dean Toste designed individual probes. Patrick R. A. Zanon and Stephan M. Hacker wrote the manuscript with input from all authors.

Profiling the proteome-wide selectivity of diverse electrophiles

Patrick R. A. Zanon[1], Fengchao Yu[2], Patricia Z. Musacchio[3], Lisa Lewald[1], Michael Zollo[1], Kristina Krauskopf[4], Dario Mrdović[1], Patrick Raunft[1], Thomas E. Maher[1], Marko Cigler[4], Christopher J. Chang[5], Kathrin Lang[4], F. Dean Toste[3], Alexey I. Nesvizhskii[2,6], and Stephan M. Hacker[1,*]

[1] Department of Chemistry, Technical University of Munich, Garching, Germany

[2] Department of Pathology, University of Michigan, Ann Arbor, Michigan, USA

[3] Department of Chemistry, University of California, Berkeley, California, USA

[4] Department of Chemistry, Group of Synthetic Biochemistry, Technical University of Munich, Garching, Germany

[5] Departments of Chemistry and Molecular and Cell Biology, University of California, Berkeley, California, USA

[6] Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, Michigan, USA

* Correspondence: stephan.m.hacker@tum.de

Abstract: Targeted covalent inhibitors are powerful entities in drug discovery, but their application has so far mainly been limited to addressing cysteine residues. The development of cysteine-directed covalent inhibitors has largely profited from determining their proteome-wide selectivity using competitive residue-specific proteomics. Several probes have recently been described to monitor other amino acids using this technology and many more electrophiles exist to modify proteins. Nevertheless, a direct, proteome-wide comparison of the selectivity of diverse probes is still entirely missing. Here, we developed a completely unbiased workflow to analyse electrophile selectivity proteome-wide and applied it to directly compare 54 alkyne probes containing diverse reactive groups. In this way, we verified and newly identified probes to monitor a total of nine different amino acids as well as the *N*-terminus proteome-wide. This selection includes the first probes to globally monitor tryptophans, histidines and arginines as well as novel tailored probes for methionines, aspartates and glutamates.

Introduction

Targeted covalent inhibitors (TCIs) are powerful entities in drug discovery as they can have key advantages such as increased binding affinity to the target, the possibility to generate selectivity among closely related proteins and better pharmacodynamic properties.[1] Nevertheless, careful optimisation of the reactivity and selectivity of these inhibitors is key to avoiding toxicity as well as possible immunogenic reactions.[1]

For TCIs that target cysteine residues, competitive residue-specific proteomics is an essential tool to tailor covalent inhibitors for high selectivity proteome-wide.[2,3] The underlying methods are based on the ground-breaking isotopic tandem orthogonal proteolysis activity-based protein profiling (isoTOP-ABPP) platform.[2,3] We have recently built on this platform by developing isotopically labelled desthiobiotin azide (isoDTB) tags to streamline the experimental procedure.[4] In the underlying isoDTB-ABPP workflow[4] (Fig. 1a), two samples of a proteome-of-interest are treated with a covalent ligand or the corresponding solvent as a control. Next, a broadly reactive alkyne probe is applied that labels many residues with alkynes. The residue(s) that are already engaged by the ligand are blocked from this reactivity and are not modified with alkynes. In the next step, isotopically differentiated isoDTB tags are attached using copper-catalysed azide-alkyne cycloaddition (CuAAC).[5] As the proteins originating from the compound- and solvent-treated samples are now differentiated, the two samples are mixed, enriched, proteolytically digested and the modified peptides eluted. Using liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS), the modified peptides are identified and quantified. Peptides containing residues that are engaged by the covalent ligand will show high ratios between the two samples (R » 1), while all unaffected peptides will show ratios close to 1. In this way, target engagement and selectivity of TCIs can be investigated proteome-wide in a quantitative fashion.[4]

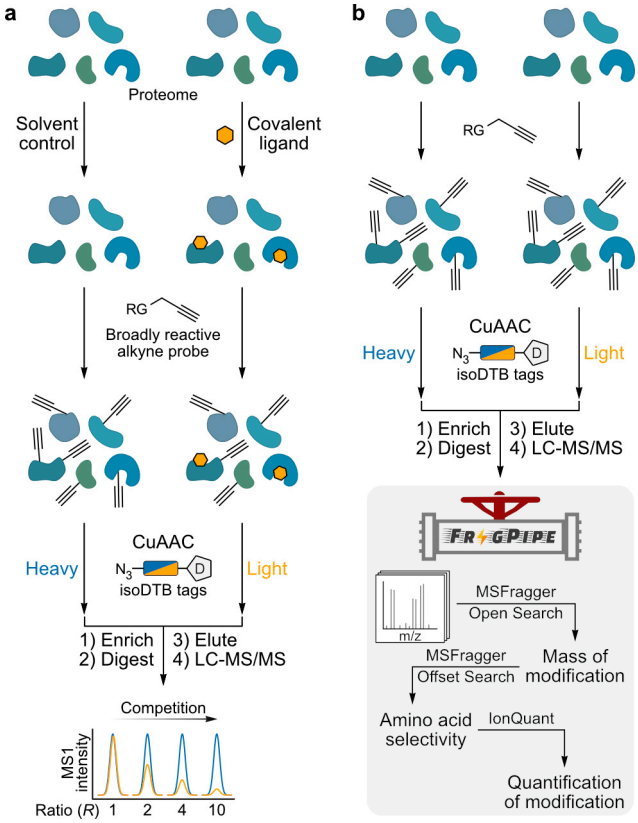


Fig. 1 | Workflows for competitive isoDTB-ABPP and the unbiased analysis of electrophile selectivity. a, Workflow for competitive, residue-specific chemoproteomic experiments using the isoDTB-ABPP workflow.[4] RG = reactive group; D = desthiobiotin. **b**, Unbiased workflow to comprehensively investigate electrophile reactivity in the proteome using the MSFragger-based FragPipe computational platform.[6,7]

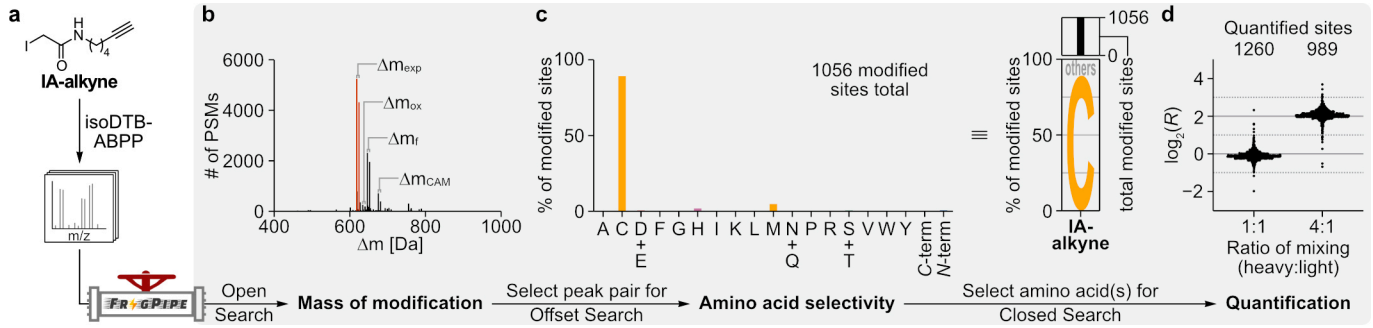


Fig. 2 | An unbiased workflow to study electrophile selectivity using the MSFragger-based FragPipe computational platform. a, Labelling of the proteome of *S. aureus* SH1000 with 1 mM **IA-alkyne** and analysis using the isoDTB-ABPP workflow (Fig 1b)[4] results in MS data for the analysis of proteome-wide reactivity and selectivity. **b**, Through analysis with an Open Search in MSFragger,[6,7] the masses of modification that occur proteome-wide are assigned. The peaks highlighted in red are the expected modifications (Δm_{exp}) resulting from alkylation and modification with the light and heavy isoDTB tags, respectively. Further modification of the alkylated peptides by oxidation (Δm_{ox}), formylation (Δm_f) or carbamidomethylation on a second cysteine (Δm_{CAM}) is also detected. PSMs = peptide spectrum matches. **c**, One peak pair (Δm_{exp}) is selected for an MSFragger Offset Search that localises this modification to the modified amino acid(s). In this way, selectivity is assessed across all proteinogenic amino acids. The bar graph represents the fraction of all modified sites that is modified at the indicated amino acid. The same data is also represented in a letter plot, in which the size of the letter is scaled by the fraction of all modified sites that is modified at the indicated amino acid. All amino acids that are modified in less than 5% of the cases are summarised as "others". The total number of modified sites is given as a bar graph on top of the letter plot. **d**, One amino acid (cysteine) is selected for quantification at the selected masses (Δm_{exp}) using an MSFragger Closed Search and the IonQuant quantification module.[8] Here, two datasets were analysed, in which the heavy and light samples were mixed at a ratio of 1:1 and 4:1, respectively. The grey, solid lines indicate the expected values of $\log_2(R) = 0$ and $\log_2(R) = 2$. The grey, dashed lines indicate the respective preferred window of quantification ($-1 < \log_2(R) < 1$ for the 1:1 ratio, $1 < \log_2(R) < 3$ for the 4:1 ratio). The total number of quantified sites is given on top of the plot. The mass spectrometric experiments used for this analysis were performed in an earlier study.[4] All data is based on technical duplicates.

Despite the promise of TCIs to address diverse reactive amino acids, they so far almost exclusively address cysteine residues.[9] However, as cysteine is a very rare amino acid, many interesting binding pockets contain no suitable cysteine for covalent engagement.[10] Recently, many reactive groups have been developed that selectively target other residues in proteins and that have promise for TCI development against a much broader set of target proteins.[11] For these compounds, one key challenge is to globally investigate their target and amino acid selectivity. Alkyne-, azide- or (desthio)biotin-containing probes have been successfully developed in various cases to directly investigate global selectivity.[12] Nevertheless, besides often being synthetically challenging, this approach requires that the modification is stable at all modified residues, which excludes this strategy *e.g.* for acylating reagents that form unstable thioesters at cysteine, which evade direct detection.[13] Similarly, this approach is not suitable for reversibly covalent inhibitors,[14,15] whose modifications would also be lost before the analysis. In contrast, if one broadly reactive alkyne probe is developed that leads to a stable modification of an amino acid of interest, the targets for all of these compounds can be identified competitively. Therefore, being able to competitively profile target engagement at various amino acids using residue-specific proteomics would enable a global profile of the selectivity of TCIs regardless of their amino acid selectivity or the stability of their adducts.

To date, broadly reactive alkyne probes that can be used in this approach have been reported for cysteines,[2,3,16,17] lysines,[18,19] aspartates and glutamates,[10,20] methionines[21,22] as well as tyrosines.[23,24] In the different underlying studies various strategies for the affinity tags, the isotopic labelling, the mass spectrometric instrumentation and the data analysis have been used. In this way, it is impossible to directly compare the reactivity and selectivity of these probes and to choose the best probe for a certain application. Furthermore, for other amino acids, broadly reactive alkyne probes are still entirely missing. While a direct comparison of the amino acid selectivity of five carbon electrophiles has been performed previously,[25] a broad scale comparison of many diverse electrophiles is still entirely missing. To address this challenge, we here directly compared a large variety of electrophiles and established their proteome-wide amino acid selectivity using isoDTB-ABPP.[4] This method is ideally suited to perform this analysis as it allows specific enrichment of the modified peptides to increase sensitivity of detection. Furthermore, it fosters confident assignment of probe-labelled peptides due to the isotopic pattern introduced by the isoDTB tags.

As we are most interested in the application of TCIs for antibacterial applications,[4] we performed the main part of our analysis in the lysate of *Staphylococcus aureus* SH1000.[26] We enhanced and developed new features for the MSFragger-based[6,7] FragPipe computational platform (https://fragpipe.nesvilab.org), which now allows studying proteome-wide electrophile reactivity in a completely unbiased fashion (Fig. 1b). In this way, we verified or newly identified probes to competitively study a total of nine different amino acids as well as the *N*-terminus proteome-wide. For the chosen amino acid-selective probes, we verify similar selectivity in the human proteome indicating that amino acid selectivity is not dependent on the specific proteome. This set of probes will enable competitive profiling of TCIs not only against cysteine but against a variety of reactive amino acids and thereby guide the development of novel TCIs.

Results and Discussion

Tailoring the MSFragger-based FragPipe computational platform for unbiased analysis of proteome-wide electrophile selectivity. One important challenge for the widespread application of competitive residue-specific proteomics is that tailored software is not readily available to the broader community.[27] Furthermore, it was so far not possible to study electrophile reactivity and selectivity in a completely unbiased fashion. In order to address these challenges, we set out to tailor the MSFragger software and other components of the FragPipe computational platform for this application.[6,7] We chose to use FragPipe as it is freely available to the academic community and as the ultrafast fragment-ion indexing method is especially powerful for the complex data analyses needed to identify and localise modifications on peptides in an unbiased fashion. To validate the new software features, we utilised a published dataset,[4] in which 1 mM iodoacetamide alkyne (**IA-alkyne**) was used in a non-competitive isoDTB-ABPP workflow (Fig. 1b).

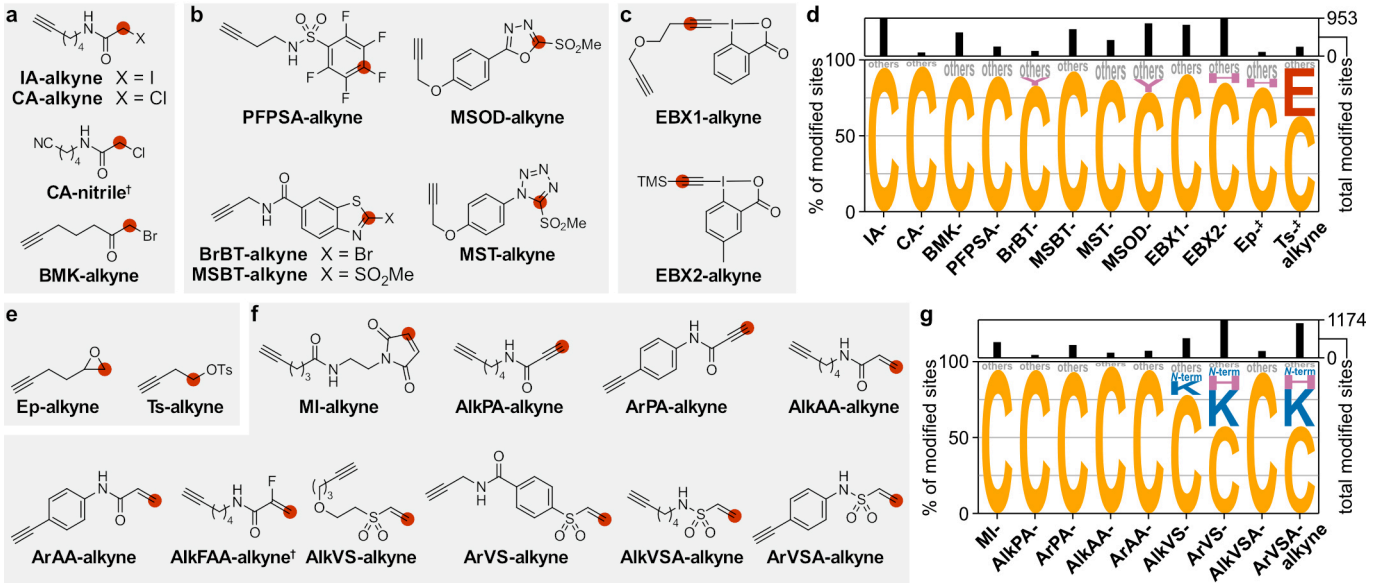


Fig. 3 | Amino acid selectivity of electrophiles targeting cysteines. a,**b**,**c**,**e**,**f**, Structures of alkyne-containing electrophilic probes that were investigated for their proteome-wide amino acid selectivity. Red circles indicate the initial site of electrophilic reactivity. **d**,**g**, Amino acid selectivity of the probes upon treatment of the proteome of *S. aureus* SH1000 at 100 μM probe concentration. The data is represented as letter plots. The size of the letter is scaled by the fraction of all modified sites that is modified at the indicated amino acid. All amino acids that are modified in less than 5% of the cases are summarised as "others". The total number of modified sites is given as a bar graph on top of the letter plot. †: No clear mass of modification was detected and therefore no analysis of the amino acid selectivity was possible. ‡: Data for the indicated probe at 1 mM is shown. All data is based on technical duplicates.

In the first step, we optimised FragPipe's Open Search workflow to identify modified peptides and assign the masses of their respective modifications. For this purpose, we performed an Open Search using MSFragger[6,7,28,29] followed by removal of mass shift artifacts with Crystal-C,[28] processing of peptide and protein identifications using PeptideProphet[30] and ProteinProphet,[31] false discovery-rate (FDR) filtering *via* Philosopher[32] and summarisation of the mass shifts observed on peptides using PTM-Shepherd.[29] Using this procedure, we identified the expected alkylation with **IA-alkyne** as the main modification (Δm_{exp}) in the form of two peaks corresponding to the isotopically differentiated isoDTB tags (Fig. 2a,b). Furthermore, we detected a significant amount of additional formylation (Δm_f) of the modified peptides, which is known to occur when eluting, re-dissolving and storing peptides in solutions containing formic acid.[33] Additionally, we identified smaller peaks that correspond to oxidation of the formed thioether (Δm_{ox}) or additional carbamidomethylation at a second cysteine residue (Δm_{CAM}). The detection of these expected minor modifications verifies that MSFragger, with its ability to perform accurate deisotoping[34] and mass calibration,[7] can identify different modifications occurring after treatment with an electrophilic probe proteome-wide in an unbiased fashion with high mass accuracy. As the highest deviation of all of these masses of modification from the expected value was 0.0044 Da (6.9 ppm), the molecular formula can be directly deduced from the MS data for unknown modifications.

After having determined the mass of the modification, we performed an MSFragger Offset Search, which allows searching at the mass with an indicated offset from the mass of the unmodified peptide.[6,7] Here, we chose an offset that accounts for the detected alkylation with **IA-alkyne** (Δm_{exp}). The modification is computationally localised to an amino acid residue or a stretch of amino acids without having to previously specify, which amino acids are potentially modified.[7] Thus, the FragPipe computational platform allows to globally study selectivity towards all amino acid residues and the protein termini at once. After thorough data filtering, we verified that **IA-alkyne** shows high selectivity of 89% for cysteines with alkylation of methionines (5%) being the most prominent additional reactivity (Fig. 2c).

Finally, to use probes for competitive residue-specific proteomics, their relative intensity between the light and the heavy channel needs to be quantified. For this purpose, complex in-house software[3,35] or workarounds in existing software[4] had to be used, so far. Therefore, we extended the quantification tool IonQuant[8] – another FragPipe component – to allow quantification of modified peptides in MSFragger Offset Searches as well as in conventional Closed Searches, in which the modified residue is specified. To do so, we implemented a new functionality in IonQuant to perform relative quantification of isotopically labelled peaks to support stable isotope labelling-based quantification. In a Closed Search, we quantified 1260 cysteines with >99% in the preferred quantification window of $-1 < \log_2(R) < 1$ (Fig. 2d). For the Offset Search, we quantified 1896 modified peptides (99% in the preferred quantification window, Supplementary Fig. 1). It should be noted that the number of total modified sites, which is used for selectivity analysis based on an Offset Search (1056 total modified sites in this case, Fig. 2c), and the number of quantified residues in the Closed Search (1260 quantified cysteines in this case, Fig. 2d) or the Offset Search (1896 quantified peptides in this case, Supplementary Fig. 1) will be different in most cases, because of different data analysis workflows and downstream data filtering. We also analysed an additional published dataset,[4] in which the heavy and light sample were mixed at a ratio of 4:1. Here, we quantified 989 cysteines with 98% in the preferred quantification window of $1 < \log_2(R) < 3$ (Fig. 2d). The high quality of this quantification data was comparable to data evaluation with MaxQuant[36] using our previously described workaround[4] or pFind 3[37] using a custom script for downstream analysis[35] (Supplementary Fig. 2). Importantly, our automated FragPipe workflow simplifies the data analysis and allows analysis for probes that are not selective for a certain amino acid type.

Overall, the optimised FragPipe computational platform enables the completely unbiased analysis of residue-specific proteomic data obtained with various probes including identification of the mass of the modification, its amino acid selectivity and its use for quantitative applications. Having this unbiased analysis workflow at hand, we applied a standardised sample preparation protocol for all probes

throughout this project to directly compare them regarding their proteome-wide reactivity and selectivity. Here, two identical samples of *S. aureus* lysate were treated with 100 μM of the respective probe, modified with 100 μM of the light or heavy isoDTB tag, respectively, mixed at a ratio of 1:1 and analysed using the isoDTB-ABPP workflow.

Diverse chemistries allow monitoring of cysteines. Using the described standard conditions for **IA-alkyne**[3] (Fig. 3a), we detected 95% selectivity for cysteines and quantified 1197 cysteines (Fig. 3d, Supplementary Fig. 3 and 4). An increase of the concentration to 1 mM resulted in 86% selectivity for cysteines. Chloroacetamide **CA-alkyne**[3] and α-bromomethyl ketone **BMK-alkyne**[38] (Fig. 3a) also demonstrated high selectivity for cysteines (96% and 89%) and allowed quantification of 230 and 976 cysteines, respectively (Fig. 3d, Supplementary Fig. 3 and 4). Importantly, a chloroacetamide negative control (**CA-nitrile**, Fig. 3a), with a nitrile replacing the alkyne, did not yield a clear mass of modification (Supplementary Fig. 3).

Nucleophilic aromatic substitution is an alternative chemistry to label cysteine residues in proteins (**PFPSA-**,[39] **BrBT-**, **MSBT-**, **MST-** and **MSOD-alkyne**,[16] Fig. 3b). All of the studied probes showed high cysteine selectivity (70-92%) and allowed quantification of 362-1061 cysteines (Fig. 3d, Supplementary Fig. 5 and 6). We also investigated hypervalent iodine reagents that have been described for labelling of cysteines in the proteome.[17] Using the originally reported chemistry (**EBX1-alkyne**, Fig. 3c, Supplementary Fig. 7),[40] we detected several different modifications (Supplementary Fig. 8).[41] While all of these showed high cysteine selectivity (91-97%, Fig. 3d, Supplementary Fig. 8 and 9), this complex modification behaviour complicates downstream data analysis. Therefore, we turned to the recently published probe **EBX2-alkyne**[42] (Fig. 3c), which selectively led to the minimal modification with an ethynyl group (Supplementary Fig. 8, 9 and 10). This modification showed high selectivity for cysteines (85%) and allowed quantification of 1251 cysteines in the proteome (Fig. 3d, Supplementary Fig. 8 and 9). Finally, we also turned to nucleophilic substitution reactions at unactivated sp^3-carbon centres (**Ep-** and **Ts-alkyne**, Fig. 3e). Both probes showed a clear preference for cysteines, but for **Ts-alkyne** a significant number of modifications was also found at glutamates (Fig. 3d, Supplementary Fig. 11 and 12).

When we combined the quantified cysteines of all these probes, we in aggregate quantified 1941 cysteines in *S. aureus* covering 37% of the 5268 cysteines encoded in the *S. aureus* genome. **IA-alkyne** can still be considered the gold standard for monitoring cysteine residues with residue-specific proteomics *in vitro*. Nevertheless, we verified that many other complementary probes exist that further increase the coverage with **EBX2-** and **MSBT-alkyne** being especially powerful (Supplementary Fig. 13).

***Michael* acceptors preferentially react with cysteines.** *Michael* acceptors are currently mainstays for the design of TCIs (Fig. 3f). For a maleimide probe (**MI-alkyne**, after hydrolysis of the resulting succinimide during the workflow) and for propiolamide probes (**AlkPA-** and **ArPA-alkyne**), we detected high cysteine selectivity (93-95%) and quantified a total of 283-752 cysteines (Fig. 3g, Supplementary Fig. 14 and 15). We also observed the expected modification and strong proteomic labelling (160-1174 localised sites, Fig. 3f,g, Supplementary Fig. 16-19) for all acceptor-substituted terminal alkenes except for **AlkFAA-alkyne**. Strikingly, we observed different degrees of selectivity. While **AlkAA-alkyne** labelled cysteine residues with 97% selectivity, less than 60% of labelled sites were cysteines for **ArVSA-** and **ArVS-alkyne** (>20% lysines, ~9% histidines, ~5% *N*-termini). Interestingly, the labelling of *N*-termini showed a preference for the secondary amine of proline when the initial methionine was removed (Supplementary Fig. 19).[43] This data demonstrates that, even though *Michael* acceptors can be designed to also target other amino acid residues, cysteines are the major modification sites in all studied cases. In this way, while lysine-selective TCIs have been developed based on *Michael* acceptors,[44] monitoring their cysteine off-targets is an important prerequisite for their further development.

Studying lysine residues proteome-wide. The activated ester probe **STP-alkyne** (Fig. 4a) allows selective monitoring of many lysine residues in the proteome.[18] Using our workflow, we verified its high selectivity for lysines (78%, Fig. 4b, Supplementary Fig. 20 and 21), which led to the quantification of 3277 lysines. The remaining peptides were mainly labelled at serines (9%), threonines (2%) or *N*-termini (5%). This selectivity was fully retained even at 1 mM probe concentration. It is noteworthy that peptides labelled at threonine preferentially have a histidine at position −2 and those labelled at serine show a preference for cysteine at position +2, arginine at position −1 and histidine at position −2 indicating that these sequence contexts might increase reactivity towards **STP-alkyne** (Supplementary Fig. 21).[43] We also studied four additional acylation reagents (**TFP-**,[18] **NHS-**,[19] **ATT-**[45] and **NASA-alkyne**,[46] Fig. 4a). and detected an overall similar selectivity, which allowed quantification of 2145-4404 lysines (Fig. 4b, Supplementary Fig. 20 and 22). All of these probes are therefore suitable to study lysines proteome-wide and hold potential especially for applications that are hampered by the negative charge of **STP-alkyne**. All of the probes showed similar preferences in the sequence around labelled serines and threonines as seen for **STP-alkyne** (Supplementary Fig. 22). As there are no tools to globally study serines and threonines in a residue-specific fashion and as tools to study the *N*-terminus lead to limited coverage (*vide infra*), it is furthermore noteworthy that **STP-alkyne** also quantified 428 serines, 165 threonines and 152 *N*-termini. This points to the fact that, while **STP-alkyne** mainly labels lysines, it could also moonlight to detect effects on these other amino acids.

Besides acylation reagents, squaric acid derivatives (**AlkSq-** and **ArSq-alkyne**, Fig. 4c) also react with amines under physiological conditions.[47] Both probes demonstrated very high selectivity for lysines (93% and 90%, Fig. 4b, Supplementary Fig. 23 and 24). Therefore, **ArSq-alkyne** is a promising broadly reactive alkyne probe for lysines (2990 quantified lysines), while structures like **AlkSq-alkyne** might be interesting for TCI design due to tempered reactivity (1339 quantified lysines). Another reactivity of lysines is the formation of imines with aldehydes, but irreversible stabilisation of the product is needed for the application as broadly reactive alkyne probe. Here, we were enticed by 2-ethynyl-benzaldehyde-based probes like **EBA-alkyne** (Fig. 4d) that were previously shown to react with amines in proteins to form an imine that cyclises to yield an irreversible isoquinolinium salt (Supplementary Fig. 25).[48] Exhibiting high lysine selectivity (81%) and good proteomic coverage (3796 quantified lysines, Fig. 4b, Supplementary Fig. 26 and 27), **EBA-alkyne** is an additional probe promising for lysine monitoring. Importantly, all of these probes also quantified a significant number of *N*-termini (243 for **AlkSq-alkyne**, 252 for **ArSq-alkyne** and 253 for **EBA-alkyne**), which demonstrates monitoring of *N*-termini as another possible application for these probes.

Light-activated probes hold the potential to monitor covalent target engagement in living cells as the probe itself is unreactive before activation and therefore has the potential to be non-toxic.[10] Proteome-wide, residue-specific monitoring of lysines in a light-dependent fashion would therefore be highly promising to study effects on lysines with temporal and spatial resolution in living cells. Nevertheless, no probe had been previously shown to allow this

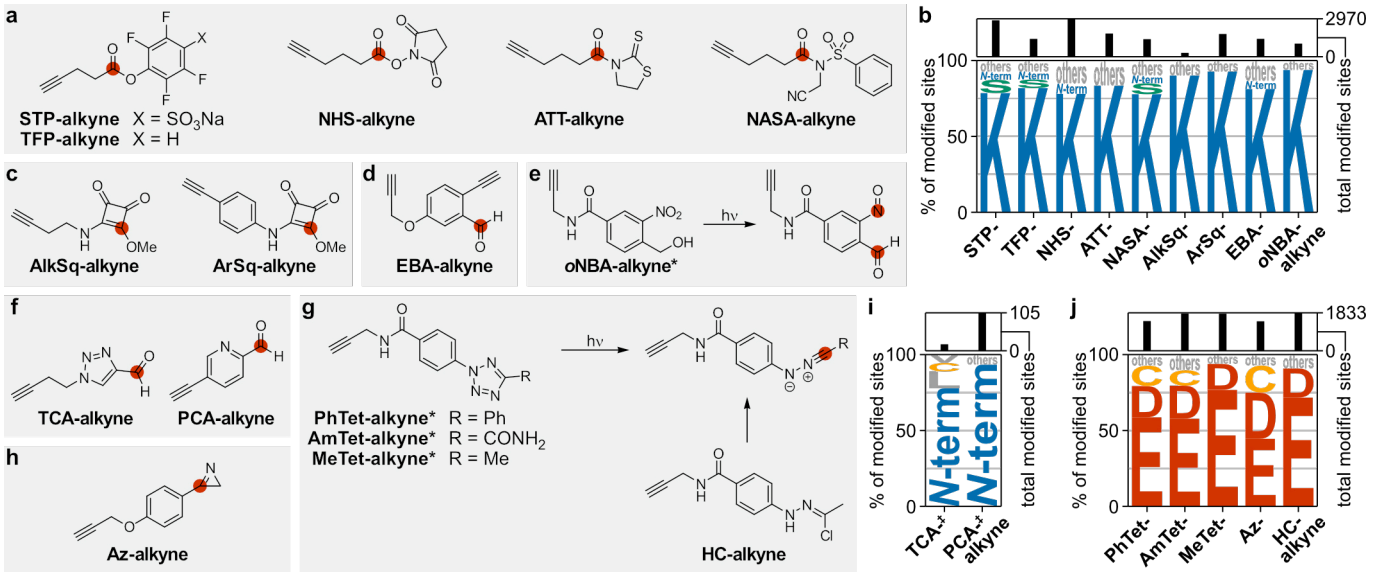


Fig. 4 | Amino acid selectivity of electrophiles targeting lysines, *N*-termini, aspartates and glutamates. a,**c**,**d**,**e**,**f**,**g**,**h**, Structures of alkyne-containing electrophilic probes that were investigated for their proteome-wide amino acid selectivity. Red circles indicate the initial site of electrophilic reactivity. For probes ***o*NBA-**, **PhTet-**, **AmTet-**, **MeTet-** and **HC-alkyne**, the reactions leading to the reactive intermediate are also shown. **b**,**i**,**j**, Amino acid selectivity of the probes upon treatment of the proteome of *S. aureus* SH1000 at 100 μM probe concentration. The data is represented as letter plots. The size of the letter is scaled by the fraction of all modified sites that is modified at the indicated amino acid. All amino acids that are modified in less than 5% of the cases are summarised as "others". The total number of modified sites is given as a bar graph on top of the letter plot. *: Labelling was performed using UV-activation at 280-315 nm (**PhTet-**, **AmTet-** and **MeTet-alkyne**) or 365 nm (***o*NBA-alkyne**) for 10 min. ‡: Data for the indicated probe at 1 mM is shown. All data is based on technical duplicates.

application. For this purpose, we set out to study ***o*NBA-alkyne** that had been shown to react with lysines in live cells after formation of a nitrosobenzaldehyde upon irradiation (Fig. 4e, Supplementary Fig. 28).[49] Indeed, we detected the expected modification (Supplementary Fig. 29 and 30) and a high degree of lysine selectivity (94%) for this probe, which was retained even at 1 mM probe concentration (93%, Fig. 4b, Supplementary Fig. 29 and 30). Therefore, this is the first demonstration that ***o*NBA-alkyne** can act as a photoprobe to globally study lysines using residue-specific proteomics (1456 quantified lysines).

Taken the data of all lysine-directed probes together, we quantified 9129 lysines, which covers 15% of the 62,166 lysines encoded in the genome of *S. aureus*. Although **STP-alkyne** remains the reagent of choice to study lysines for standard applications, **ArSq-**, **EBA-** and ***o*NBA-alkyne** also display high selectivity using complementary chemistries. These probes will be important not only for increasing overall coverage (Supplementary Fig. 31), but also to allow specific applications, especially within living cells.

Global monitoring of *N*-termini of proteins. Although several lysine-directed probes like **STP-**, **ArSq-** and **EBA-alkyne** also allow monitoring of *N*-termini, selective chemistry is highly desirable. Here, carboxaldehydes of electron-poor heteroaromatics have been previously applied to modify proteins (**TCA-**[50] and **PCA-alkyne**,[51] Fig. 4f). While we were not able to detect more than a few sites for **TCA-alkyne** in the proteome even at 1 mM, **PCA-alkyne** resulted in the expected modification by initial formation of an imine with the *N*-terminus followed by cyclisation involving the first amide nitrogen of the backbone (Supplementary Fig. 32 and 33). This modification showed high selectivity for the *N*-terminus (93%) and allowed quantification of 167 *N*-termini at 1 mM probe concentration (Fig. 4i, Supplementary Fig. 33 and 34). Even though the coverage certainly still needs to be improved, **PCA-alkyne** is a suitable starting point to selectively monitor the *N*-terminus proteome-wide. Taking the data of **PCA-** (1 mM), **STP-**, **ArSq-** and **EBA-alkyne** together, we were able to quantify 464 *N*-termini in 412 proteins (some proteins were detected with and without clipping of the *N*-terminal methionine) covering 14% of the 2959 proteins encoded in the *S. aureus* genome.

Monitoring aspartates and glutamates in the proteome. Recently, we have developed 2,5-disubstituted tetrazoles (Fig. 4g) into light-activatable tools to globally study aspartates and glutamates in living bacteria.[10] Using our workflow, we verified the expected modification by formation of a nitrilimine upon light irradiation and subsequent reactivity eventually leading to an diacylated hydrazine (**PhTet-**, **AmTet-** and **MeTet-alkyne**, Fig. 4g, Supplementary Fig. 35 and 36). This modification showed a strong preference for aspartates and glutamates (79-94%, Fig. 4j, Supplementary Fig. 36-38). The best probe, **MeTet-alkyne**, quantified 2192 aspartates and glutamates and its high selectivity was retained at 1 mM probe concentration (93%, Supplementary Fig. 36 and 38). While peptides labelled at aspartate were consistently enriched for aliphatic amino acids at position +1 (leucine, valine, isoleucine), those labelled at glutamate showed a preference for valine and alanine at positions −1 and −2 (Supplementary Fig. 37 and 38).[43] When prolonging the photoreaction with **MeTet-alkyne**, we observed a decrease in the expected modification and an increasing additional modification that could be assigned as arylation, which localised to all aromatic amino acids and to cysteines (Supplementary Fig. 36 and 38). While this chemistry might therefore also be interesting to monitor aromatic amino acids, care must be taken to not overextend irradiation times when monitoring aspartates and glutamates with **MeTet-alkyne**.

2*H*-Azirines have been described as constitutively active electrophiles to target aspartates and glutamates by initial nucleophilic attack at the azirine eventually leading to an acylated α-aminoketone (**Az-alkyne**,[20] Fig. 4h, Supplementary Fig. 39). Using our workflow, we could show that besides the expected reactivity (Δm_{exp}), these probes also showed a mass of modification with a mass of Δm_{exp}+1 Da, which formally corresponds to hydrolysis leading to a loss of ammonia (Supplementary Fig. 39 and 40). As computationally differentiating these modifications was not fully possible, we further analysed their combined selectivity (Fig. 4j,

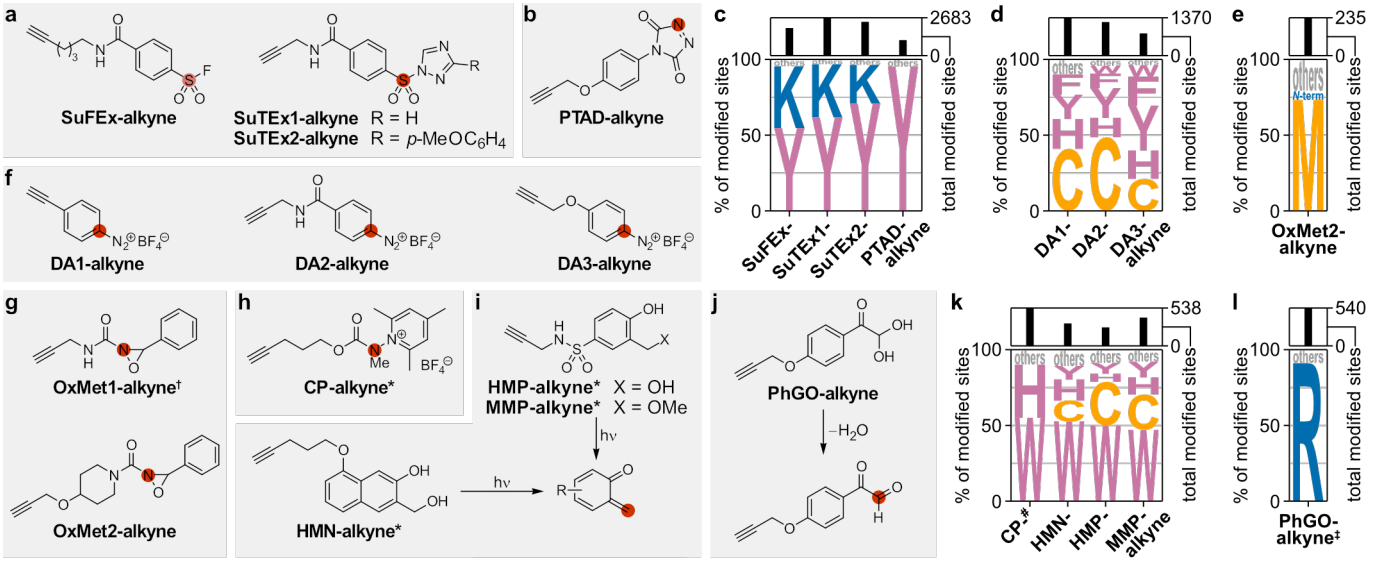


Fig. 5 | Amino acid selectivity of electrophiles targeting tyrosines, aromatic amino acids, methionines, tryptophans, histidines and arginines. a,**b**,**f**,**g**,**h**,**i**,**j**, Structures of alkyne-containing electrophilic probes that were investigated for their proteome-wide amino acid selectivity. Red circles indicate the initial site of electrophilic reactivity. For probes **HMN-**, **HMP-**, **MMP-** and **PhGO-alkyne**, the reactions leading to the reactive intermediate are also shown. **c**,**d**,**e**,**k**,**l**, Amino acid selectivity of the probes upon treatment of the proteome of *S. aureus* SH1000 at 100 μM probe concentration. The data is represented as letter plots. The size of the letter is scaled by the fraction of all modified sites that is modified at the indicated amino acid. All amino acids that are modified in less than 5% of the cases are summarised as "others". The total number of modified sites is given as a bar graph on top of the letter plot. †: No clear mass of modification was detected and therefore no analysis of the amino acid selectivity was possible. *: Labelling was performed using UV-activation at 280-315 nm (**CP-**, **HMP-** and **MMP-alkyne**) or 365 nm (**HMN-alkyne**) for 10 min. ‡: Data for the indicated probe at 1 mM is shown. #: Labelling was performed in degassed lysate under argon. All data is based on technical duplicates.

Supplementary Fig. 40 and 41) and identified 75% of all modifications at aspartates and glutamates and 18% at cysteines. **Az-alkyne** is therefore a valuable probe to study aspartates and glutamates, but care must be taken to account for cysteine off-targets.

Enticed by the high selectivity of **MeTet-alkyne**, we also explored similar chemistry for monitoring aspartates and glutamates with constitutively active electrophiles. Hydrazonoyl chlorides such as **HC-alkyne** (Fig. 4g) have been shown to liberate the same reactive nitrilimine as the light-activated tetrazoles (Supplementary Fig. 35).[52] We detected the expected modification and high selectivity for aspartates and glutamates (91%), which was also retained at 1 mM probe concentration (92%, Fig. 4j, Supplementary Fig. 42 and 43). The sequence around the labelled amino acids showed the same preferences as also seen for **MeTet-alkyne** (Supplementary Fig. 43).[43] **HC-alkyne** allowed the quantification of 2450 aspartates and glutamates in the proteome at 100 μM and is therefore a novel promising probe to study these amino acids without the need for light-activation.

Interestingly, **MeTet-** and **HC-alkyne** showed a strong preference to label glutamates (77% and 71%) over aspartates (17% and 19%), while **Az-alkyne** demonstrated a smaller difference (44% and 30%, respectively) indicating that it is better able to also react with the sterically more hindered aspartate. Taking all carboxylic acid-directed probes together, 7811 aspartates and glutamates were quantified corresponding to 7.8% of the 100,780 aspartates and glutamates encoded in the *S. aureus* genome. Specifically, **MeTet-**, **HC-** and **Az-alkyne** constitute a set of complementary probes that allow a deep profiling of these amino acids in the bacterial proteome (Supplementary Fig. 44). Considering that no probes exist to selectively monitor *C*-termini proteome-wide, it is furthermore noteworthy that all carboxylic acid-directed probes together quantified 179 *C*-termini with **MeTet-** and **HC-alkyne** at 1 mM being most promising (101 and 109 quantified *C*-termini).

Residue-specific proteomics at tyrosines. Tyrosines offer a unique opportunity for various selective chemistries through reactions with the hydroxyl group as well as with the electron-rich aromatic system. Recently, sulfonylation of the hydroxyl group using sulfur-fluoride (**SuFEx-alkyne**)[53] and sulfur-triazole exchange chemistry (**SuTEx1-** and **SuTEx2-alkyne**)[23,24] has been established for proteome-wide approaches (Fig. 5a). We verified the tyrosine reactivity of these probes (55-71%, Fig. 5c, Supplementary Fig. 45 and 46) with lysine residues being the most prominent off-targets (26-41%). In agreement with a previous study in human proteomes, **SuTEx2-alkyne** showed the highest tyrosine selectivity and allowed quantifying 2653 tyrosines in bacterial lysates.

For labelling of the aromatic system of tyrosines in proteins, reagents like **PTAD-alkyne** have been established (Fig. 5b, Supplementary Fig. 47).[54] We identified the expected adduct that shows high selectivity for tyrosines (95%, Fig. 5c, Supplementary Fig. 48 and 49). Furthermore, we also detected a modification that corresponds to fragmentation of **PTAD-alkyne** to the isocyanate and subsequent reactivity, which showed some selectivity for lysines and *N*-termini and could be strongly reduced through addition of excess primary amine (Supplementary Fig. 48 and 49).[55] Due to the exquisite selectivity of the expected modification, it would be interesting to optimise the stability of these reagents to further reduce this side reactivity.

Through combination of the data for all tyrosine probes, we quantified 3968 tyrosines covering 12% of the 32,172 tyrosines encoded in the *S. aureus* genome. While **SuTEx2-alkyne** is currently the probe of choice for tyrosines, reagents like **PTAD-alkyne** should allow for the development of optimised complementary probes in the future (Supplementary Fig. 50).

Diazonium salts perform arylation chemistry in the bacterial proteome. During the investigation of tyrosine-directed chemistries, we also considered aryl diazonium salts that have been shown to lead to azo coupling on tyrosines in isolated proteins (Fig. 5f, Supplementary Fig. 51).[56] Proteome-wide, we only detected minor

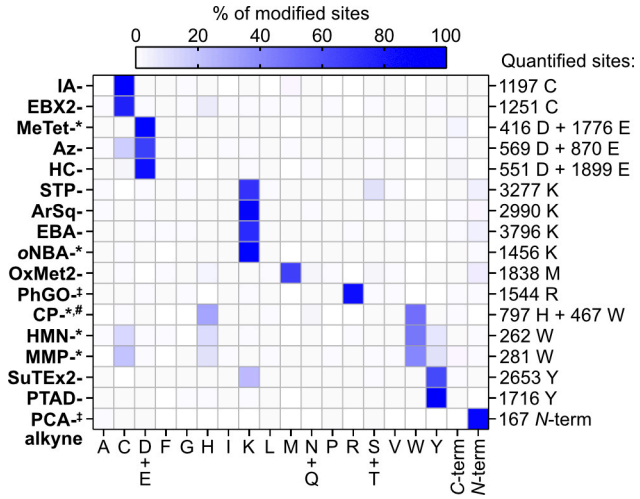


Fig. 6 | A set of 17 electrophilic probes that enables studying nine different amino acid residues and the *N*-terminus. For a selection of the probes that allow studying diverse residues in the proteome of *S. aureus*, their selectivity is plotted in a heatmap. The colour is scaled by the fraction of all modified sites that is modified at the indicated amino acid. *: Labelling was performed using UV-activation at 280-315 nm (**MeTet-**, **CP-** and **MMP-alkyne**) or 365 nm (**oNBA-** and **HMN-alkyne**) for 10 min. ‡: Data for the indicated probe at 1 mM is shown. #: Labelling was performed in degassed lysate under argon. All data is based on technical duplicates.

azo coupling and almost exclusively arylation, corresponding to a formal loss of molecular nitrogen (Supplementary Fig. 51-53).[57] Strikingly, next to modifications on cysteines,[58] this also led to up to 75% of all modifications being localised to aromatic amino acids for **DA3-alkyne** (19% phenylalanines, 19% histidines, 8% tryptophans, 30% tyrosines, Fig. 5d, Supplementary Fig. 52 and 53). In this way, **DA3-alkyne** enabled the quantification of 1218 aromatic amino acids. While aryl diazonium salts are therefore not suitable for azo couplings in bacterial proteomes, it is noteworthy that they can be used to monitor various aromatic amino acids and that preferred reactivity is possible at many residues in the proteome including even the unactivated aromatic system of phenylalanines.

An optimised tool for global monitoring of methionine residues. Hypervalent iodine reagents have been described to monitor methionines but require an additional reaction step to give a stable modification.[22] Therefore, we focused on oxaziridines for labelling of methionines, which proceeds through initial attack of the thioether at the nitrogen of the oxaziridine eventually giving a sulfimide (Supplementary Fig. 54).[21] The initially described reagent **OxMet1-alkyne** (Fig. 5g) did not result in detection of the expected modification using our standard workflow (Supplementary Fig. 55). Implementing recently described design principles for more stable methionine modification,[59] we synthesised **OxMet2-alkyne** (Fig. 5g), which led to the detection of a high number of modified peptides with a preference for methionine modification (73%, 1838 quantified methionines, 8.5% of the 21,677 methionines encoded in the *S. aureus* genome, Fig. 5e, Supplementary Fig. 55 and 56). In this way, we developed **OxMet2-alkyne** as a tailored reagent for proteome-wide monitoring of methionines.

The first technologies to monitor tryptophans and histidines proteome-wide. Recently, *N*-carbamoylpyridinium salts like **CP-alkyne** (Fig. 5h) have been introduced to photochemically label tryptophans in proteins through photoinduced electron transfer leading to the attachment of the carbamoyl group to tryptophans (Supplementary Fig. 57).[60] However, proteome-wide studies were still missing. Upon irradiation under protective gas, **CP-alkyne** showed the expected mass of modification in the proteome with almost complete selectivity for tryptophans (55%) and histidines (35%, Fig. 5k, Supplementary Fig. 58 and 59) allowing the quantification of 467 tryptophans and 797 histidines. The selectivity was retained, when the reaction was run open to air or at 1 mM **CP-alkyne** (Supplementary Fig. 58 and 59). While we could not detect any preferences in the sequence around labelled histidine, peptides modified at tryptophan consistently showed an enrichment of glutamate in the −1-position (Supplementary Fig. 59).[43] The reactivity with histidines is especially noteworthy, as we were unsuccessful to detect the expected modification with reported histidine-selective thiophosphorodichloridates (**TPAC-alkyne**, Supplementary Fig. 60), most likely due to instability of the conjugate during our workflow.[61] The fraction of >35% of histidine labelling for **CP-alkyne** is the highest we have detected for any probe, making it the probe of choice to study histidines proteome-wide. Taking all conditions for **CP-alkyne** together, we quantified 1697 histidines covering 9.1% of the 18,746 histidines in the *S. aureus* genome. **CP-alkyne** is, therefore, the first probe to globally monitor both tryptophans and histidines using residue-specific proteomics.

Simultaneously, we also investigated UV-activatable *o*-quinone methide precursors for tryptophan labelling through a formal [4+2]-cycloaddition (**HMN-**, **HMP-** and **MMP-alkyne**, Fig. 5i, Supplementary Fig. 61). Also in this case, we observed a preference for tryptophans (47-53%, Fig. 5k, Supplementary Fig. 62 and 63) with cysteines (14-29%), histidines (5-13%) and tyrosines (7-10%) being the main off-targets. In this way, **HMN-** and **MMP-alkyne** are complementary probes to study tryptophans in a proteome-wide sense. Through the combination of the data for all probes, we quantified a total of 701 tryptophans covering 11% of the 6183 tryptophans encoded in the *S. aureus* genome (Supplementary Fig. 64).

A first-in-class method to monitor arginines in the whole proteome. Although arginine has been targeted with glyoxal-based reagents for bioconjugation[62] and crosslinking before,[63] no method is currently available to globally monitor arginines with residue-specific proteomics. We therefore synthesised **PhGO-alkyne** based on the known reactivity of phenylglyoxals with arginines that eventually produces a stable imidazole derivative (Fig. 5j, Supplementary Fig. 65). Using our workflow, we detected only minor modification of the proteome with the expected modification accompanied by an oxidation product (Supplementary Fig. 66).[64] At increased concentration of 1 mM **PhGO-alkyne**, however, we detected many modifications of the proteome (1544 arginines quantified, 5.4% of the 28,550 arginines encoded in the *S. aureus* genome) with the expected mass and high arginine selectivity (91%) alongside some oxidation (Fig. 5l, Supplementary Fig. 66 and 67). While the presence of the oxidation product still necessitates some optimisation, **PhGO-alkyne** is the first probe to globally monitor arginine in the proteome.

A set of probes to globally study nine amino acids and the *N*-terminus. Through screening of 54 alkyne probes, we have identified a set of 17 probes that we currently consider ideal for profiling nine different amino acids and the *N*-terminus proteome-wide (Fig. 6). The fraction of all residues of a certain amino acid that we quantified using this probe set ranged from 2.5% of all aspartates to 29% of all cysteines encoded in the genome of *S. aureus*. In total, we were able to quantify 20,558 different sites in the proteome using our probe selection. These sites cover 1399 of 2959 proteins encoded in the *S. aureus* genome (54%) and 85% of the annotated essential proteins (301/353).[65] In this way, our probe selection allows us to gain very deep insights into the bacterial proteome.

To show that the application of our probe selection is not restricted to bacterial systems, we also applied these probes in lysates of the human cancer cell line MDA-MB-231. Except for **PCA-alkyne**, we reproduced similar selectivities for all probes in the human proteome (Supplementary Fig. 68-76). In this way, we were able to quantify a total of 23,831 sites in 3770 proteins in this human cell line.

Conclusion

We report here on a universal, unbiased workflow to study the reactivity and selectivity of electrophilic probes in a proteome-wide setup. For this purpose, we extended and established new components of the broadly available MSFragger-based FragPipe computational platform.[6,7] These include an optimised workflow for filtering of Open Searches to reliably identify masses of modification, a new workflow for downstream data analysis of Offset Searches to investigate selectivity at all proteogenic amino acids simultaneously and a new feature for IonQuant that now allows to quantify data from stable isotopic labelling-based MS experiments. In this way, we are for the first time able to identify the added masses of modification, their selectivity at all proteinogenic amino acids and their relative abundance in a completely unbiased fashion.

While our selection of 54 alkyne probes is certainly not comprehensive, we are convinced that we covered most chemistries that have more broadly been applied for protein labelling and that do not require additional reagents. In this way, we verified and newly identified tailored probes to study nine different amino acids as well as the *N*-terminus proteome-wide. Within our set of probes, we verify the selectivity of **IA-** and **EBX2-alkyne** for cysteines,[2,3] **STP-alkyne** for lysines,[18] **SuTEx2-alkyne** for tyrosines[23] as well as **MeTet-**[10] and **Az-alkyne**[20] for aspartates and glutamates. For several of these amino acids, we identify additional probes based on complementary chemistries like **ArSq-** and **EBA-alkyne** for lysines, **HC-alkyne** for aspartates and glutamates and **PTAD-alkyne** for tyrosines, which increase the proteome-wide coverage. For lysines, we demonstrate for the first time that proteome-wide residue-specific profiling is possible in a light-dependent fashion using **oNBA-alkyne**, which opens up exciting opportunities for in cell studies. Additionally, we developed a tailored probe (**OxMet2-alkyne**) for methionines that allows application in our workflow with high coverage due to increased stability of the protein modification. Furthermore, we transferred the concept of labelling proteins at the *N*-terminus to global residue-specific studies using **PCA-alkyne**. Strikingly, our study has enabled us to develop probes that for the first time allow studying tryptophans (**CP-**, **HMN-**, **MMP-alkyne**), histidines (**CP-alkyne**) and arginines (**PhGO-alkyne**) in a residue-specific fashion in the whole proteome.

Important residues, for which no selective broadly reactive alkyne probes could be established here, are serines and threonines as well as *C*-termini. For serines and threonines, bioconjugation to proteins using P(V)-based reagents for phosphorus-sulfur incorporation (PSI) has recently been established.[66,67] Nevertheless, we were unable to detect a clear modification for **PSI-alkyne** (Supplementary Fig. 77 and 78), most likely due to instability of the conjugate during our workflow. In our hands, **STP-alkyne** is still the best way to monitor at least some serines and threonines in a residue-specific fashion. For *C*-termini, the aspartate- and glutamate-directed probes allowed us to quantify 179 *C*-termini, but increased coverage and selective probes are surely desirable here.

One important application of our probe selection will be to competitively study target engagement for any covalently reactive protein ligand regardless of its amino acid selectivity and adduct stability. This will be especially important for ligands, for which probe synthesis is challenging or which form covalent adducts that evade direct detection like those forming unstable modifications such as thioesters[13] or covalently reversible protein adducts.[14,15] It is tempting to speculate that in the future it will be possible to use a mixture of several probes to competitively profile the selectivity of a protein ligand simultaneously at various amino acids within the same proteomic experiment as has recently been described for the combined detection of cysteines and lysines.[13]

Taking all of this together, we report on an unbiased, universal workflow to characterise amino acid selectivity of electrophilic compounds in the whole proteome. These studies have enabled the profiling of established, tailored as well as new probes for residue-specific proteomics, which now allows monitoring of a total of nine different amino acids and the *N*-terminus. We are convinced that our workflow and this probe selection will be instrumental to identifying selectively reactive groups for the design of covalent inhibitors targeting diverse amino acids. In this way, they will advance the development of TCIs for the many protein binding sites that lack a suitable cysteine residue for engagement with the currently dominant acrylamide-based chemistry.

Data availability

Mass spectrometric data for all proteomic analyses have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) *via* the PRIDE partner repository[68] with the dataset identifier PXD024454. Raw data for Fig. 2-6 is available in Supplementary Tables 1-4. All other data is available in the main text or the Supplementary Information.

Code Availability

Code for the MSFragger-based FragPipe analyses is available at https://fragpipe.nesvilab.org/.

Acknowledgements

S.M.H. and P.R.A.Z. acknowledge funding by the Fonds der Chemischen Industrie through a Liebig Fellowship and a Ph.D. fellowship and by the TUM Junior Fellow Fund. A.I.N. acknowledges financial support from the US National Institutes of Health (R01-GM-094231 and U24-CA210967). F.D.T. thanks the Novartis Institutes for BioMedical Research and the Novartis-Berkeley Center for Proteomics and Chemistry Technologies for supporting this work. P.Z.M. thanks the American Cancer Society for a post-doctoral fellowship (PF-18-132-01-CDD). M.Z. gratefully acknowledges funding by the Studienstiftung des Deutschen Volkes through a Ph.D. fellowship. K.L. is grateful for support by the collaborative research centre SFB1035 (German Research Foundation DFG, Sonderforschungsbereich 1035, Projektnummer 201302640, project B10). C.J.C. acknowledges funding by the NIH (ES 28096 and GM139245). We gratefully acknowledge Stephan A. Sieber and his group for their generous support. We thank the group of Jérôme Waser (EPFL) for providing **EBX1-alkyne**, Ku-Lung Hsu and Jeffrey W. Brulet (University of Virginia) for providing **SuTEx1-alkyne** and **SuTEX2-alkyne** and the group of Phil S. Baran (Scripps Research) for providing **PSI-alkyne**. We thank Simon Kollmannsberger and Maximilian Iglhaut (Technical University of Munich) for assisting with the synthesis of probes as well as Katja Bäuml and Mona Wolff (Technical University of Munich) for technical assistance. We thank Daniel Geiszler, Felipe da Veiga Leprevost and Dmitry Avtonomov for useful discussions and technical assistance regarding new developments in FragPipe.

Author Contributions

P.R.A.Z. and S.M.H. designed the research, planned experiments, performed and analysed proteomics experiments. F.Y. and A.I.N. designed, developed and benchmarked data analysis software. P.R.A.Z. synthesised the majority of probes. P.Z.M., L.L., M.Z., K.K., D.M., P.R., T.E.M., and M.C. contributed the synthesis of individual probes. K.L., C.J.C. and F.D.T. designed individual probes. P.R.A.Z. and S.M.H. wrote the manuscript with input from all authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper in the online version.

Correspondence and requests for materials should be addressed to S.M.H.

References

1 Singh, J., Petter, R. C., Baillie, T. A. & Whitty, A. The resurgence of covalent drugs. *Nat. Rev. Drug. Discov.* **10**, 307-317 (2011).

2 Backus, K. M. *et al.* Proteome-wide covalent ligand discovery in native biological systems. *Nature* **534**, 570-574 (2016).

3 Weerapana, E. *et al.* Quantitative reactivity profiling predicts functional cysteines in proteomes. *Nature* **468**, 790-795 (2010).

4 Zanon, P. R. A., Lewald, L. & Hacker, S. M. Isotopically Labeled Desthiobiotin Azide (isoDTB) Tags Enable Global Profiling of the Bacterial Cysteinome. *Angew. Chem. Int. Ed.*, 2829-2836 (2020).

5 Rostovtsev, V. V., Green, L. G., Fokin, V. V. & Sharpless, K. B. A stepwise huisgen cycloaddition process: copper(I)-catalyzed regioselective "ligation" of azides and terminal alkynes. *Angew. Chem. Int. Ed.* **41**, 2596-2599 (2002).

6 Kong, A. T., Leprevost, F. V., Avtonomov, D. M., Mellacheruvu, D. & Nesvizhskii, A. I. MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. *Nat. Methods* **14**, 513-520 (2017).

7 Yu, F. *et al.* Identification of modified peptides using localization-aware open search. *Nat. Commun.* **11**, 4065 (2020).

8 Yu, F. *et al.* Fast Quantitative Analysis of timsTOF PASEF Data with MSFragger and IonQuant. *Mol. Cell.Proteom.* **19**, 1575-1585 (2020).

9 Sutanto, F., Konstantinidou, M. & Dömling, A. Covalent inhibitors: a rational approach to drug discovery. *RSC Med. Chem.* **11**, 876-884 (2020).

10 Bach, K., Beerkens, B. L. H., Zanon, P. R. A. & Hacker, S. M. Light-Activatable, 2,5-Disubstituted Tetrazoles for the Proteome-Wide Profiling of Aspartates and Glutamates in Living Bacteria. *ACS Cent. Sci.* **6**, 546-554 (2020).

11 Gehringer, M. & Laufer, S. A. Emerging and Re-Emerging Warheads for Targeted Covalent Inhibitors: Applications in Medicinal Chemistry and Chemical Biology. *J. Med. Chem.* **62**, 5673-5724 (2019).

12 Parker, C. G. & Pratt, M. R. Click Chemistry in Proteomic Investigations. *Cell* **180**, 605-632 (2020).

13 Cao, J. *et al.* Multiplexed CuAAC Suzuki–Miyaura Labeling for Tandem Activity-Based Chemoproteomic Profiling. *Anal. Chem.* **93**, 2610–2618 (2021).

14 Lee, K. M., Le, P., Sieber, S. A. & Hacker, S. M. Degrasyn exhibits antibiotic activity against multi-resistant Staphylococcus aureus by modifying several essential cysteines. *Chem. Commun.* **56**, 2929-2932 (2020).

15 Senkane, K. *et al.* The Proteome-Wide Potential for Reversible Covalency at Cysteine. *Angew. Chem. Int. Ed.* **58**, 11385-11389 (2019).

16 Motiwala, H. F., Kuo, Y.-H., Stinger, B. L., Palfey, B. A. & Martin, B. R. Tunable Heteroaromatic Sulfones Enhance in-Cell Cysteine Profiling. *J. Am. Chem. Soc.* **142**, 1801-1810 (2020).

17 Abegg, D. *et al.* Proteome-Wide Profiling of Targets of Cysteine reactive Small Molecules by Using Ethynyl Benziodoxolone Reagents. *Angew. Chem. Int. Ed.* **54**, 10852-10857 (2015).

18 Hacker, S. M. *et al.* Global profiling of lysine reactivity and ligandability in the human proteome. *Nat. Chem.* **9**, 1181-1190 (2017).

19 Ward, C. C., Kleinman, J. I. & Nomura, D. K. NHS-Esters As Versatile Reactivity-Based Probes for Mapping Proteome-Wide Ligandable Hotspots. *ACS Chem. Biol.* **12**, 1478-1483 (2017).

20 Ma, N. *et al.* 2H-Azirine-Based Reagents for Chemoselective Bioconjugation at Carboxyl Residues Inside Live Cells. *J. Am. Chem. Soc.* **142**, 6051-6059 (2020).

21 Lin, S. *et al.* Redox-based reagents for chemoselective methionine bioconjugation. *Science* **355**, 597-602 (2017).

22 Taylor, M. T., Nelson, J. E., Suero, M. G. & Gaunt, M. J. A protein functionalization platform based on selective reactions at methionine residues. *Nature* **562**, 563-568 (2018).

23 Hahm, H. S. *et al.* Global targeting of functional tyrosines using sulfur-triazole exchange chemistry. *Nat. Chem. Biol.* **16**, 150-159 (2019).

24 Brulet, J. W., Borne, A. L., Yuan, K., Libby, A. H. & Hsu, K.-L. Liganding Functional Tyrosine Sites on Proteins Using Sulfur–Triazole Exchange Chemistry. *J. Am. Chem. Soc.* **142**, 8270-8280 (2020).

25 Weerapana, E., Simon, G. M. & Cravatt, B. F. Disparate proteome reactivity profiles of carbon electrophiles. *Nat. Chem. Biol.* **4**, 405-407 (2008).

26 Horsburgh, M. J. *et al.* sigmaB modulates virulence determinant expression and stress resistance: characterization of a functional rsbU strain derived from Staphylococcus aureus 8325-4. *J. Bacteriol.* **184**, 5457-5467 (2002).

27 Benns, H. J., Wincott, C. J., Tate, E. W. & Child, M. A. Activity- and reactivity-based proteomics: Recent technological advances and applications in drug discovery. *Curr. Opin. Chem. Biol.* **60**, 20-29 (2021).

28 Chang, H.-Y. *et al.* Crystal-C: A Computational Tool for Refinement of Open Search Results. *J. Proteome Res.* **19**, 2511-2515 (2020).

29 Geiszler, D. J. *et al.* PTM-Shepherd: analysis and summarization of post-translational and chemical modifications from open search results. *Mol. Cell. Proteom.* (2020).

30 Keller, A., Nesvizhskii, A. I., Kolker, E. & Aebersold, R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.* **74**, 5383-5392 (2002).

31 Nesvizhskii, A. I., Keller, A., Kolker, E. & Aebersold, R. A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.* **75**, 4646-4658 (2003).

32 da Veiga Leprevost, F. *et al.* Philosopher: a versatile toolkit for shotgun proteomics data analysis. *Nat. Methods* **17**, 869-870 (2020).

33 Lenčo, J., Khalikova, M. A. & Švec, F. Dissolving Peptides in 0.1% Formic Acid Brings Risk of Artificial Formylation. *J. Proteome Res.* **19**, 993-999 (2020).

34 Teo, G. C., Polasky, D. A., Yu, F. & Nesvizhskii, A. I. Fast Deisotoping Algorithm and Its Implementation in the MSFragger Search Engine. *J. Proteome Res.* **20**, 498-505 (2021).

35 Fu, L. *et al.* A quantitative thiol reactivity profiling platform to analyze redox and electrophile reactive cysteine proteomes. *Nat. Protoc.* **15**, 2891-2919 (2020).

36 Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367-1372 (2008).

37 Chi, H. *et al.* Comprehensive identification of peptides in tandem mass spectra using an efficient open search engine. *Nat. Biotechnol.* **36**, 1059-1061 (2018).

38 Abo, M. & Weerapana, E. A Caged Electrophilic Probe for Global Analysis of Cysteine Reactivity in Living Cells. *J. Am. Chem. Soc.* **137**, 7087-7090 (2015).

39 Embaby, A. M., Schoffelen, S., Kofoed, C., Meldal, M. & Diness, F. Rational Tuning of Fluorobenzene Probes for Cysteine-Selective Protein Modification. *Angew. Chem. Int. Ed.* **57**, 8022-8026 (2018).

40 Frei, R. *et al.* Fast and Highly Chemoselective Alkynylation of Thiols with Hypervalent Iodine Reagents Enabled through a Low Energy Barrier Concerted Mechanism. *J. Am. Chem. Soc.* **136**, 16563-16573 (2014).

41 Tessier, R. *et al.* "Doubly Orthogonal" Labeling of Peptides and Proteins. *Chem* **5**, 2243-2263 (2019).

42 Tessier, R. *et al.* Ethynylation of Cysteine Residues: From Peptides to Proteins in Vitro and in Living Cells. *Angew. Chem. Int. Ed.* **59**, 10961-10970 (2020).

43 O'Shea, J. P. *et al.* pLogo: a probabilistic approach to visualizing sequence motifs. *Nat. Methods.* **10**, 1211-1212 (2013).

44 Pettinger, J., Jones, K. & Cheeseman, M. D. Lysine-Targeting Covalent Inhibitors. *Angew. Chem. Int. Ed.* **56**, 15200-15209 (2017).

45 Hirata, T. *et al.* Synthesis and reactivities of 3-indocyanine-green-acyl-1,3-thiazolidine-2-thione (ICG-ATT) as a new near-infrared fluorescent-labeling reagent. *Bioorg. Med. Chem.* **6**, 2179-2184 (1998).

46 Tamura, T. *et al.* Rapid labelling and covalent inhibition of intracellular native proteins using ligand-directed N-acyl-N-alkyl sulfonamide. *Nat. Commun.* **9**, 1870 (2018).

47 Ivancová, I., Pohl, R., Hubálek, M. & Hocek, M. Squaramate-Modified Nucleotides and DNA for Specific Cross-Linking with Lysine-Containing Peptides and Proteins. *Angew. Chem. Int. Ed.* **58**, 13345-13348 (2019).

48 Deng, J.-R. *et al.* N-Terminal selective modification of peptides and proteins using 2-ethynylbenzaldehydes. *Commun. Chem.* **3**, 67 (2020).

49 Guo, A.-D. *et al.* Light-induced primary amines and o-nitrobenzyl alcohols cyclization as a versatile photoclick reaction for modular conjugation. *Nat. Commun.* **11**, 5472 (2020).

50 Onoda, A., Inoue, N., Sumiyoshi, E. & Hayashi, T. Triazolecarbaldehyde Reagents for One-Step N-Terminal Protein Modification. *ChemBioChem* **21**, 1274-1278 (2020).

51 MacDonald, J. I., Munch, H. K., Moore, T. & Francis, M. B. One-step site-specific modification of native proteins with 2-pyridinecarboxyaldehydes. *Nat. Chem. Bio.* **11**, 326-331 (2015).

52 Zengeya, T. T. *et al.* Co-opting a Bioorthogonal Reaction for Oncometabolite Detection. *J. Am. Chem. Soc.* **138**, 15813-15816 (2016).

53 Dong, J., Krasnova, L., Finn, M. G. & Sharpless, K. B. Sulfur(VI) Fluoride Exchange (SuFEx): Another Good Reaction for Click Chemistry. *Angew. Chem. Int. Ed.* **53**, 9430-9448 (2014).

54 Ban, H., Gavrilyuk, J. & Barbas, C. F. Tyrosine bioconjugation through aqueous ene-type reactions: a click-like reaction for tyrosine. *J. Am. Chem. Soc.* **132**, 1523-1525 (2010).

55 Ban, H. *et al.* Facile and Stabile Linkages through Tyrosine: Bioconjugation Strategies with the Tyrosine-Click Reaction. *Bioconjug. Chem.* **24**, 520-532 (2013).

56 Zhang, J., Ma, D., Du, D., Xi, Z. & Yi, L. An efficient reagent for covalent introduction of alkynes into proteins. *Org. Biomol. Chem.* **12**, 9528-9531 (2014).

57 Nothling, M. D. *et al.* Bacterial Redox Potential Powers Controlled Radical Polymerization. *J. Am. Chem. Soc.* **143**, 286-293 (2021).

58 Naveen, N., Sengupta, S. & Chandrasekaran, S. Metal-Free S-Arylation of Cysteine Using Arenediazonium Salts. *J. Org. Chem.* **83**, 3562-3569 (2018).

59 Christian, A. H. *et al.* A Physical Organic Approach to Tuning Reagents for Selective and Stable Methionine Bioconjugation. *J. Am. Chem. Soc.* **141**, 12657-12662 (2019).

60 Tower, S. J., Hetcher, W. J., Myers, T. E., Kuehl, N. J. & Taylor, M. T. Selective Modification of Tryptophan Residues in Peptides and Proteins Using a Biomimetic Electron Transfer Process. *J. Am. Chem. Soc.* **142**, 9112-9118 (2020).

61 Jia, S., He, D. & Chang, C. J. Bioinspired Thiophosphorodichloridate Reagents for Chemoselective Histidine Bioconjugation. *J. Am. Chem. Soc.* **141**, 7294-7301 (2019).

62 Dovgan, I. *et al.* Arginine-selective bioconjugation with 4-azidophenyl glyoxal: application to the single and dual functionalisation of native antibodies. *Org. Biomol. Chem.* **16**, 1305-1311 (2018).

63 Jones, A. X. *et al.* Improving mass spectrometry analysis of protein structures with arginine-selective chemical cross-linkers. *Nat. Commun.* **10**, 3911 (2019).

64 Thompson, D. A., Ng, R. & Dawson, P. E. Arginine selective reagents for ligation to peptides and proteins. *J. Pept. Sci.* **22**, 311-319 (2016).

65 Chaudhuri, R. R. *et al.* Comprehensive identification of essential Staphylococcus aureus genes using Transposon-Mediated Differential Hybridisation (TMDH). *BMC Genomics* **10**, 291 (2009).

66 Vantourout, J. C. *et al.* Serine-Selective Bioconjugation. *J. Am. Chem. Soc.* **142**, 17236-17242 (2020).

67 Knouse, K. W. *et al.* Unlocking P(V): Reagents for chiral phosphorothioate synthesis. *Science* **361**, 1234-1238 (2018).

68 Perez-Riverol, Y. *et al.* The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res.* **47**, D442-d450 (2019).

III. Conclusion

The Potential of Antibiotic TCIs

The rapid emergence and spread of bacterial resistance is a growing global threat to human health that urgently requires new advances in the research and development of antibiotics.[1] While they have been avoided in rational drug development for a long time, advantages like increased and longer-lasting potency have rekindled the interest in covalent drugs.[2] Moreover, covalent modes of action are prevalent in successful antibacterial drugs of natural origin like fosfomycin[3] and penicillins.[4] In order to discover novel antibiotic targets and lead compounds alike, inverse drug discovery is a highly promising approach and involves the entire proteome rather than only isolated proteins with known biological function.[5] In this approach, even relatively small libraries of covalent fragments can cover large fractions of chemical space and are thus ideally suited for this drug development strategy.[6] Competitive residue-specific proteomics has emerged as a means to determine the proteome-wide target profile of unmodified covalent ligands and investigate their modes of action.[7,8]

Profiling the Bacterial Cysteinome with isoDTB-ABPP

The described new chemoproteomic platform isoDTB-ABPP is ideally suited to streamline the discovery of antibiotic TCIs with new modes of action. The isoDTB tags are rapidly synthesized using standard solid-phase peptide synthesis and are compatible with a wide variety of chemical conditions as well as alternative proteases. These are advantages over the TEV-protease cleavable linkers used in the original isoTOP-ABPP platform[9] and since a second enzymatic digestion is unnecessary, the experimental workflow is shortened. In a direct comparison, the coverage of the bacterial cysteinome with **IA-alkyne** as probe is 27% higher for isoDTB-ABPP. In accordance with previous studies in the human proteome,[9] we found that cysteines of high reactivity are more likely to be located in functional sites in the proteome of *S. aureus*. By screening a small commercial covalent fragment library with over 200 members for antibiotic activity and selecting the most potent compounds for analysis in competitive isoDTB-ABPP experiments, we created a cysteine ligandability map. The distinct target profiles of the covalent fragments of various reactivities indicate a strong influence of the non-covalent binding elements on the interaction with the proteins. Finally, we validated one of the most promising screening hits by complementary *in vitro* experiments with the recombinantly expressed protein.

Profiling Aspartates and Glutamates

Aspartates and glutamates constitute a relatively large fraction of the bacterial proteome[10] and are underexplored residues for TCIs. We analyzed three 2,5-disubstituted tetrazole probes that form nitrilimines upon UV irradiation[11] with our isoDTB-ABPP platform in the lysate of *S. aureus* and quantified ~8000 aspartates and glutamates to undergo reaction with this reactive intermediate. For these probes, methyl substitution of the formed electrophilic carbon gave the highest selectivity (85%). We investigated the underlying mechanism for this selectivity on a small-molecule level and found that the cysteine adduct is photolabile and that little to no reaction occurs with other nucleophilic amino acids. We introduced a hydrazonoyl chloride as a new kind of aspartate and glutamate-directed covalent ligand and used it, along with two isoxazolium salts,[12] in competitive experiments. While we were able to detect and validate some liganded residues, the overall weak proteome-wide competition highlights the need for optimized electrophilic motifs. Due to their lack of toxicity, the photoactivatable probes additionally allowed the profiling of aspartates and glutamates in living bacteria, including Gram-negative strains. During the preparation of the manuscript, azirines were reported to target carboxylic acid residues with high selectivity[13] and thus present complementary, constitutively active electrophiles.

Profiling the Proteome-Wide Selectivity of Electrophiles

In order to assess and compare the reactivity and selectivity of many different electrophiles that have been described to engage nucleophilic amino acid side chains,[14] we developed a universal workflow with unbiased data analysis. In the first step, all masses of the modifications occurring on the peptides are analyzed, which allows for the verification of the expected and the detection of unexpected reactivities. The modifications are then localized to the residues they occurred on for amino acid selectivity analysis, followed by their quantification. We confirmed the selectivities of previously reported broadly reactive probes and introduced the first probes to monitor tryptophans, histidines, and arginines proteome-wide. Additionally, we described tailored probes for methionines, aspartates, and glutamates. Overall, we have identified 17 electrophilic probes for the profiling of nine different amino acid residues and the *N*-terminus across the proteome. The insights gained on the selectivity of these and other electrophiles will foster the development of TCIs for different nucleophilic residues and thus contribute to the expansion of accessible target space.

Opportunities and Future Directions

Most studies described here have been carried out in the proteome of *S. aureus*. Extending our isoDTB platform and the set of amino acid-selective probes to other prioritized bacteria, particularly Gram-negative strains, will provide many starting points for the development of antibiotics.

While reversible covalent binding can be advantageous *e.g.* for reduction of off-target occupancy,[15] it can hinder direct target deconvolution. Competitive workflows, however, can facilitate target identification.[15] Similarly, the detection of unstable irreversible modifications like thioesters should be possible in a competitive setting. Here, the possibility to now study nine amino acids competitively will largely help to establish target selectivities.

Even though not all covalent ligands perturb the physiological functions of their target proteins, they can serve as starting points for the development of bifunctional molecules that *e.g.* induce target degradation by recruitment of proteases.[16]

We envision that the isoDTB platform together with the identified set of selective, broadly reactive probes as well as the optimized FragPipe data analysis software suite will thus be valuable tools also for a variety of applications. The simplified experimental and computational workflows will thus hopefully make chemoproteomic experiments applicable also in groups that are not specialized in this field of research.

Several recent innovations facilitate an increased coverage of the proteome with broadly reactive probes. MS instrumentation enabling *e.g.* ion mobility spectrometry[17], new data acquisition schemes,[18] optimized sample-preparation workflows,[17] and multiplexing enabled by tandem mass tags[19] or complementary bioorthogonal chemistries[20] are remarkable new advances in the field. The potential combination of these methods with the probes developed in this work holds great potential to enlarge the target space that can be monitored.

References

[1] World Health Organization, *2020 Antibacterial Agents in Clinical and Preclinical Development:* an overview and analysis, **2021**.

[2] J. Singh, R. C. Petter, T. A. Baillie, A. Whitty, The resurgence of covalent drugs, *Nat. Rev. Drug Discov.* **2011**, *10*, 307–317.

[3] M. E. Falagas, E. K. Vouloumanou, G. Samonis, K. Z. Vardakas, Fosfomycin, *Clin. Microbiol. Rev.* **2016**, *29*, 321–347.

[4] A. L. Demain, R. P. Elander, The beta-lactam antibiotics, *Antonie van Leeuwenhoek* **1999**, *75*, 5–19.

[5] D. E. Mortenson *et al.*, "Inverse Drug Discovery" Strategy To Identify Proteins That Are Targeted by Latent Electrophiles As Exemplified by Aryl Fluorosulfates, *J. Am. Chem. Soc.* **2018**, *140*, 200–210.

[6] W. Lu *et al.*, Fragment-based covalent ligand discovery, *RSC Chem. Biol.* **2021**, *2*, 354–367.

[7] K. M. Backus *et al.*, Proteome-wide covalent ligand discovery in native biological systems, *Nature* **2016**, *534*, 570–574.

[8] S. M. Hacker *et al.*, Global profiling of lysine reactivity and ligandability in the human proteome, *Nat. Chem.* **2017**, *9*, 1181–1190.

[9] E. Weerapana *et al.*, Quantitative reactivity profiling predicts functional cysteines in proteomes, *Nature* **2010**, *468*, 790–795.

[10] UniProt: a hub for protein information, *Nucleic Acids Res.* **2015**, *43*, D204-12.

[11] S. Zhao *et al.*, Photo-induced coupling reactions of tetrazoles with carboxylic acids in aqueous solution: application in protein labelling, *Chem. Commun.* **2016**, *52*, 4702–4705.

[12] P. Martín-Gago *et al.*, Covalent Protein Labeling at Glutamic Acids, *Cell Chem. Biol.* **2017**, *24*, 589-597.e5.

[13] N. Ma *et al.*, 2H-Azirine-Based Reagents for Chemoselective Bioconjugation at Carboxyl Residues Inside Live Cells, *J. Am. Chem. Soc.* **2020**, *142*, 6051–6059.

[14] M. Gehringer, S. A. Laufer, Emerging and Re-Emerging Warheads for Targeted Covalent Inhibitors, *J. Med. Chem.* **2019**, *62*, 5673–5724.

[15] K. Senkane *et al.*, The Proteome-Wide Potential for Reversible Covalency at Cysteine, *Angew. Chem. Int. Ed.* **2019**, *58*, 11385–11389.

[16] H. Kiely-Collins, G. E. Winter, G. J. L. Bernardes, The role of reversible and irreversible covalent chemistry in targeted protein degradation, *Cell Chem. Biol.* **2021**.

[17] T. Yan *et al.*, SP3-FAIMS Chemoproteomics for High-Coverage Profiling of the Human Cysteinome, *ChemBioChem* **2021**.

[18] I. A. Hendriks, V. Akimov, B. Blagoev, M. L. Nielsen, MaxQuant.Live Enables Enhanced Selectivity and Identification of Peptides Modified by Endogenous SUMO and Ubiquitin, *J. Proteome. Res.* **2021**, *20*, 2042–2055.

[19] M. Kuljanin *et al.*, Reimagining high-throughput profiling of reactive cysteines for cell-based screening of large electrophile libraries, *Nat. Biotechnol.* **2021**, *39*, 630–641.

[20] J. Cao *et al.*, Multiplexed CuAAC Suzuki-Miyaura Labeling for Tandem Activity-Based Chemoproteomic Profiling, *Anal. Chem.* **2021**, *93*, 2610–2618.